
GENERATIVE MODELS FOR VISION AND MEDICAL IMAGING: EIGHT COORDINATES, 27 SUBTASKS, AND AN EVIDENCE-BASED RESEARCH AGENDA

Fengming Lin

ABSTRACT

Generative-model names mix operations, conditions, data regimes and representations. We organize visual generation into eight composable coordinates and 27 selected subitems, explaining each through machine learning, computer vision and medical imaging. A fixed-query arXiv frame contains 25,056 records from 2018 through 2026 year-to-date; a separate purposive set contains 60 source families inspected at recorded depths. Twenty representative publications per subitem supply 540 timeline positions and 383 distinct publications, not 540 full-text reviews. After presenting the complete taxonomy, we synthesize lexical trajectories, cross-coordinate co-mentions and source-specific validation. The evidence separates growing interest in action-conditioned generation from demonstrated action validity, joint-output consistency from independent truth, and large synthetic cohorts from target-population support. Selected medical workflows expose the importance of measurement fidelity, diagnostic structures, treatment quantities and validated simulation endpoints. Four medical-oriented opportunities are then formulated around selective decision reliability, independently checked joint dynamics, tail-risk virtual cohorts and action-grounded observation models. Each specifies existing neighbors, a necessary comparator, an endpoint and a stopping condition. The agenda favors a verified downstream improvement over changing model families or following keyword volume.

Keywords Generative modeling · Computer vision · Medical imaging · Taxonomy · Evidence synthesis

1 Scope and Methods

The object of this survey is visual generation, not every use of generative AI. We connect machine-learning formulations to CV tasks and medical-image applications, including synthesis, translation, editing, restoration, measurement reconstruction and temporal generation. Text appears as a condition or paired output; general language, audio and code generation are outside the central scope. Existing vision and medical surveys and the structural reference offer complementary coverage [1–3]. Our purpose is to explain what each task means and what evidence would make a next study useful, rather than to catalogue model names or forecast acceptance at a venue.

1.1 Discovery, Quantitative Frame, and Evidence Selection

Three collections must not be conflated. The multi-source discovery archive contains 128,499 source records, 124,215 conservatively deduplicated families and 58,516 automatic candidates; these are not manually included papers. The quantitative analysis instead uses one frozen visual-task arXiv query: 18,037 unique IDs from 2018–2025 in 35 hash-checked pages and 7,019 from 2026 in eight further pages, totaling $N = 25,056$. The snapshot finished on 21 October 2026 at 15:00 UTC; returned 2026 publication dates end on 20 October. Category and phrase restrictions, indexing lag and arXiv participation limit coverage. Separately, 60 purposively selected source families support the qualitative evidence pass, spanning ML foundations, CV methods, medical neighbors, clinical studies and guidance. None of these collections represents complete journal or conference coverage.

Inspection depth remains explicit. Selected full-text methods, results or limitations passages were checked for 54 of the 60 families. Two reviews received scope inspection, the user-provided reference structural inspection, one clinical

MRI source indexed primary-section inspection, one FDA source official-scope inspection, and one staining neighbor abstract-only inspection. This is not cover-to-cover reading of every source. Targeted searches combined task synonyms with validation, cohort, motion, clinical and control terms, followed by direct-neighbor and contrary-evidence checks. Versions and retrieval hashes remain separate even when arXiv/journal aliases are merged. Additional correspondence examples have their own source/depth log; the extraction has one Codex-assisted reviewer, not independent clinical adjudication.

1.2 Coding, Trajectories, and Sensitivity

The relevance audit argues against a filtered-corpus field share. The historical 102-rule system retains 12,237 and rejects 5,800 of 18,037 records. With seed 20261021, 24 titles and complete abstracts were inspected in each early/late period by retained/rejected stratum. Relevant counts were 19/24 and 20/24 among retained records, and 6/24 and 12/24 among rejected records, with one uncertain early rejection. These conditional counts reveal both contamination and missed work; they are not a gold-standard accuracy estimate. Sensitivity compares the full query frame, rule-retained records, title-only coding and exclusion of review titles for complete years 2018–2025. No separate 2026 relevance audit or per-label precision/recall validation is claimed.

Figure trajectories measure weighted vocabulary, while headings count matches. Let b_{ij} denote a binary title–abstract match for concept j . Freeze historical weights as

$$d_j = 1 + \ln \frac{18,037 + 1}{\text{df}_{j,2018:2025} + 1}, \quad z_i = \frac{b_i d}{\|b_i d\|_2}, \quad I_{jy} = \frac{1}{N_y} \sum_{i: \text{year}(i)=y} z_{ij}, \quad (1)$$

with $z_i = 0$ for unmatched records. All 102 concepts participate in normalization, not only the 27 displayed ones. Applying the same IDF to 2026 preserves historical weights. Annual denominators include all queried records, not just matches. Dashed terminal segments and open markers indicate 2026 YTD, without annualization. Co-mention analysis in Section 4 reports its own conditional denominators; it is not method adoption, causal dependence or a matched-calendar growth estimate.

1.3 Representative Histories and Reading Order

The histories establish landmarks, not a global citation Top 20. Each of the 27 subitems contains 20 different representative papers chosen for methodological relevance, capability coverage and recent developments: 540 memberships, 383 distinct publications. These are metadata/title–abstract checks, not 383 additional full-text reviews. Dates use verified arXiv v1 submissions where available, with exceptions logged individually; an earlier project or OpenReview release is not ruled out. Timelines cover 2018–2026*, through 22 October. Papers are equally spaced chronologically and use black arrows with the repeated height pattern $-1, +1, -2, +2$; year intervals are deliberately nonuniform and do not encode elapsed duration. Earlier foundations remain in the text. Shared papers resolve to one bibliography entry.

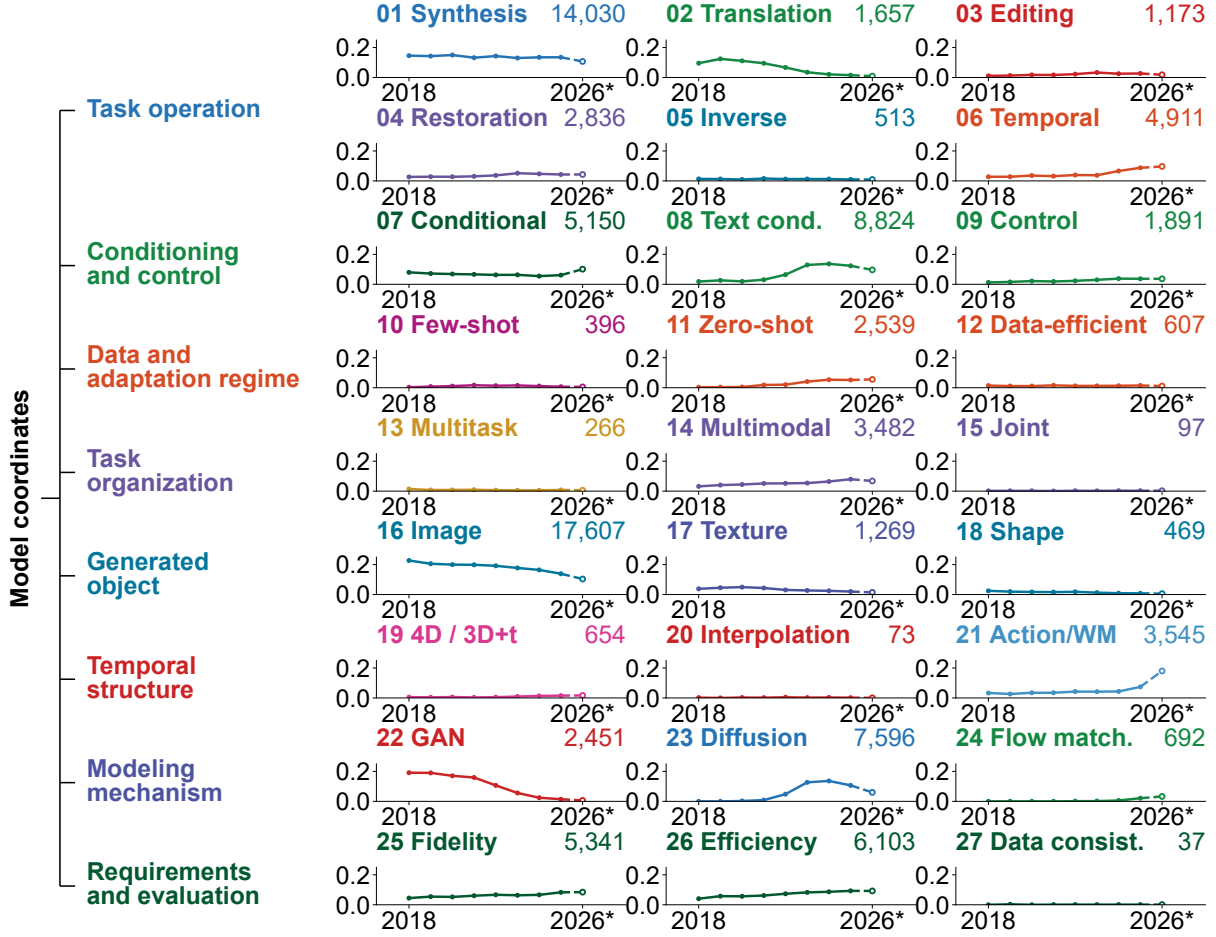
The presentation proceeds from the complete eight-coordinate overview to all 27 explanations and histories in coordinate order. Only then do we synthesize cross-dimensional evidence, examine selected medical needs and derive research opportunities. Correspondence diagrams describe shared formulations and illustrative applications, not a measured ML-to-clinic transfer pipeline. Their Need/Test statements specify evaluation requirements, not achievements attributed to every example.

2 Overview: Eight Coordinates and 27 Subitems

Read a model name as a combination of coordinates. Figure 1 aligns eight first-level groups with 27 numbered subitems. For example, few-shot text-conditioned texture synthesis combines a data regime, condition, object and operation; these are not four competing tasks. The atlas retains one pooled trajectory per panel and introduces the vocabulary expanded in the following section.

Generative-model coordinates and subtask trends

N = 25,056 | Headings: text-match counts | Curves: mean concept TF-IDF



2026*: YTD to 21 Oct. Shared y: 0-0.25. Fixed 2018-2025 IDF.

Figure 1: **Generative-model coordinates and subtask trajectories.** Eight groups, 27 selected subitems. Counts are overlapping title–abstract matches in $N = 25,056$ records, including 7,019 from 2026 YTD. Curves use Equation 1 and common 0–0.25 limits. Asterisks, open endpoints and dashed segments mark incomplete 2026, retrieved 21 October. WM denotes world model.

The first-level groups ask different questions. Operation specifies what changes; conditioning specifies information supplied; data regime specifies the target-data budget; organization specifies relationships between tasks or outputs; object specifies what is generated; temporal structure specifies dependence across time or action; mechanism specifies how a distribution is learned or sampled; requirements specify what must be preserved and tested. The 27 leaves are selected vocabulary anchors, not an exhaustive partition. A paper can occupy several leaves and most useful experiments cross several coordinates. Table 1 makes the interpretation contract explicit without repeating the atlas counts.

The coordinate contract identifies which assumption changes when two methods are compared. Its final column names common category errors that otherwise make apparent progress difficult to interpret.

Table 1: **Eight-coordinate reading contract.** Scope: the survey’s complete taxonomy; groups are composable, not exclusive.

Coordinate	Question to answer	Do not infer
Task operation	What is created, changed or recovered?	Every plausible sample recovers an observed case.
Conditioning and control	Which input is supplied, and what can be steered?	Conditioning guarantees selective control.
Data and adaptation regime	Which real examples and pretrained resources are allowed?	Few target samples mean little total training.
Task organization	What shares learning or forms a coupled sample?	Multimodal, multitask and joint are synonyms.
Generated object	Intensity, appearance or geometry?	A 3D intensity volume is a surface model.
Temporal structure	What depends on time, history or action?	Smoothness establishes valid dynamics.
Modeling mechanism	Which objective, probability path and sampler?	Backbone choice defines the task.
Requirements and evaluation	Which quantity, budget and failure must be checked?	An attractive mean score certifies every use.

Distributional and measurement conditions imply different obligations. Unconditional sampling models $p_\theta(x)$ and semantic conditioning models $p_\theta(x | c)$. A particular measurement instead constrains a particular object:

$$y = A(x) + \epsilon, \quad p(x | y, A) \propto p(y | x, A)p(x). \quad (2)$$

A prompt admits many suitable images; patient data refer to one underlying case. The operator, noise and prior all influence a reconstruction. A deterministic encoder–decoder need not define a sampling distribution, while a variational latent model does so explicitly [4]. Restricting solutions to a generator’s range can also introduce representation error [5]. Probability notation describes the intended target; it does not certify a practical sampler.

A coordinate must not stand in for its missing neighbors. Few-shot target adaptation may depend on abundant source pretraining [6]. Texture may mean surface appearance on supplied geometry, tissue appearance, or a quantitative imaging signature. A Transformer is a backbone, not by itself a distributional objective; a dynamic output is not necessarily an action model. These distinctions prevent a change of representation or terminology from being mistaken for a new task. The next section follows items 01–27 exactly, with a correspondence map before each group’s individual histories.

3 Eight Dimensions in Detail

Each group begins with a shared ML–CV–medical map, followed by its numbered subitems. The maps supply examples and invariants; the histories supply dated landmarks. All 27 items are introduced here before any cross-dimensional synthesis or research opportunity.

3.1 Task operation

An operation says what is created, changed or recovered. The six rows separate new-instance sampling from identity-preserving translation, local edits, missing-content recovery, measurement inversion and sequence generation. Their conditions and acceptable errors differ even when one model family supports several operations.

Figure 2 gives the examples for this coordinate of Figure 1.

01 Task operation

Figure 1 coordinate 1 | Subitems 01, 02, 03, 04, 05, 06

Machine learning

01 Synthesis

Noise + condition -> new sample
Learn a sample distribution
Keep requested properties
Test coverage and diversity

02 Translation

Source domain -> target domain
Learn a conditional mapping
Keep relevant subject identity
Pairing changes assumptions

03 Editing

Image + edit -> modified image
Change a specified property
Keep non-target content
Not a treatment counterfactual

04 Restoration / fill

Degraded input -> restored output
Recover or fill missing content
Keep observed information
Separate truth from completion

05 Measurement inverse

Data + operator -> reconstruction
Infer a constrained posterior
Unobserved parts stay uncertain
Noise assumptions matter

06 Temporal generation

Past / action -> future states
Model dependent temporal states
Keep meaningful dynamics
Actions are optional conditions

Computer vision

— Image / texture / shape

Noise / prompt -> visual sample
LDM; Text2Tex
Keep requested attributes
Test fidelity and coverage

— Image-to-image translation

Source -> target appearance
pix2pix; CycleGAN
Keep scene / object structure
Test subject-wise preservation

— Instruction editing

Image + instruction -> edit
InstructPix2Pix
Keep identity outside the edit
Test target / off-target change

— Denoise / enhance / inpaint

Partial image -> restored image
Palette; DDPM
Keep supported detail
Test false structures

— Inverse reconstruction

Known operator -> image draws
Diffusion posterior sampling
Keep likelihood-supported detail
Test operator shift / calibration

— Video / 4D / world models

History / control -> future view
Video diffusion; Genie; 4D-fy
Keep temporal / action relations
Test unseen actions

Medical image analysis

— MAISI synthetic CT

Masks / metadata -> CT volume
Vary appearance, retain anatomy
Need rare-case and privacy QA
Test held-out real-task utility

— Synthetic CT / virtual stain

MRI -> CT; tissue -> stain
MR-only RT; virtual histology
Need subject-specific fidelity
Test dose / pathology and QA

— Anatomic digital siblings

Local edit -> related anatomy
Diffusion-based local editing
Need controlled variation
Test local truth and invariants

— Microscopy / region filling

Degraded to enhanced content
Microscopy; MAISI tumor fill
Separate recovery / simulation
Test hallucination + task error

— MRI / sparse-view CT

k-space / projections -> image
Score-based medical inversion
Need patient / lesion fidelity
Test data shifts / decision error

— Cardiac cycles / ultrasound

Phase / pose -> anatomy / view
CHeart; Mesh Flow; SonoWorld
Need local motion / action truth
Test against static / replay

Figure 2: **Task operation: ML–CV–medical correspondences.** Subitems 01, 02, 03, 04, 05, 06. Meanings and illustrative uses; Need/Test labels are requirements, not evidence that every example passed. Sources: [7–26].

3.1.1 01 Synthesis

Synthesis samples an instance rather than recovers a specific observed person. A text-to-image model creates a scene compatible with a description; a medical volume generator creates a possible anatomy under supplied conditions. The input can be noise alone or a combination of noise and metadata, while the output may be pixels, voxels, geometry or texture. Latent diffusion is a representative visual implementation [7]. In medicine, MAISI conditions volumetric generation on body region and spacing and provides mask-guided or tumor-filling paths [9]. Texture is another object coordinate: Text2Tex generates surface appearance on an already supplied mesh, so it does not demonstrate joint geometry recovery [8]. The elementary question is which attributes the conditions control and which variations remain free, not simply whether the rendered result looks realistic.

Synthesis is useful when its variation supports a real downstream need. Training augmentation requires improvement on held-out real cases, not performance on generated labels alone; cohort construction requires representative coverage, not sample count alone. MAISI includes external autoPET image-distribution evaluation, but that is not an independent diagnostic endpoint or privacy audit [9]. MAISI-v2 combines rectified flow, region-specific losses and changed data/training conditions, so differences cannot all be attributed to one objective [27]. Conditional anatomy and rare congenital topology also have existing implementations [28, 29]. A responsible comparison measures conditional accuracy, tail support, memorization risk and real-task usefulness at matched resources. If resampling or a statistical shape model already supplies the required variation, a larger learned generator must justify its extra assumptions and costs.

Figure 3 places 20 representative synthesis publications in the common 2018–2026* window; the arrows resolve to the bibliography [7, 30–48]. The common selection and date conventions are defined in Section 1.

01 Synthesis

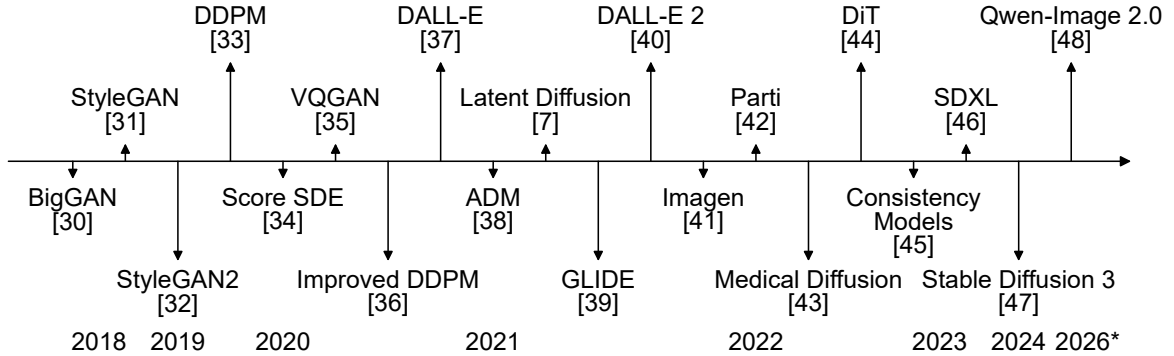


Figure 3: **01 Synthesis: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.1.2 02 Translation

Translation changes the domain while trying to preserve the relevant object. Turning a label map into a photograph changes appearance; turning MRI into synthetic CT changes an imaging representation; virtual staining maps an unstained measurement to stain-like contrast. Paired translation learns from aligned examples, whereas unpaired translation learns from marginal domain samples with additional constraints [10, 11]. The distinction is consequential: cycle consistency can support an invertible mapping without uniquely identifying the medically correct subject-wise mapping. Virtual staining demonstrates a genuine medical translation task, with registration and tissue preparation affecting the reference pairs [13]. The output should therefore be described as a conditional estimate under a specific acquisition and training protocol, not a new measurement of every biological quantity that the target modality might reveal.

The preservation contract determines the evaluation. For synthetic CT, density and geometric errors matter through dose calculation and positioning, while for staining they matter through structures and diagnostic observations. AortaDiff already couples contrast translation with segmentation and clinically interpretable measurements, so “task-aware joint translation” is not an untouched idea [49]. Its preprint evaluation also has ROI, alignment and derived-reference limitations that should remain visible. The SynthRAD2025 report separates image, geometric and dose criteria and demonstrates why rankings depend on the endpoint [50]. Unpaired realism or average similarity is insufficient when a rare structure changes treatment planning. Translation should be compared against the current clinical workflow, including non-generative alternatives, and failures should trigger an explicit fallback rather than be hidden in a favorable mean score.

Figure 4 places 20 representative translation publications in the common 2018–2026* window; the arrows resolve to the bibliography [51–70]. The common selection and date conventions are defined in Section 1.

02 Translation

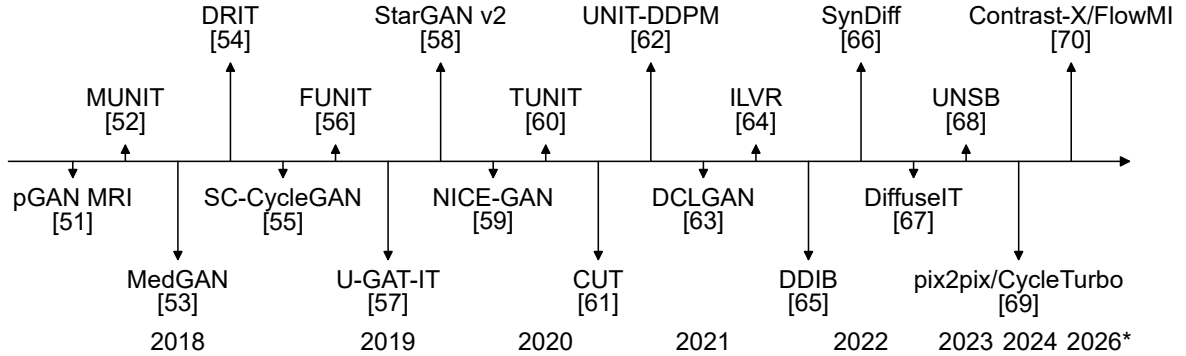


Figure 4: **02 Translation: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.1.3 03 Editing

Editing asks for a controlled difference from an existing instance. An instruction such as changing an object’s color is not unconditional sampling; the original image supplies identity and non-target attributes. InstructPix2Pix learns from generated instruction-edit triplets and reports difficulties with unwanted changes, isolation and viewpoints [14]. In medicine, digital siblings already create local anatomic variations and assess morphology and topology [15]. Here, the useful output is often a family of controlled alternatives for sensitivity analysis rather than a clinically asserted future patient. One must specify the editable region, required invariants, permitted topology changes and the meanings of edit magnitude. A visually small change can still be large for a downstream measurement, especially near a lesion, valve or vessel boundary.

A counterfactual image is not automatically a causal treatment counterfactual. An edited anatomy can support a computational experiment if the simulator and intervention are explicitly defined, but visual plausibility alone does not identify how a patient would respond to treatment. Distributional editing also cannot establish that all non-target physiology is preserved. Appropriate tests compare independent measurements before and after editing, stress boundary cases, and quantify collateral changes outside the requested region. Classical parameterized deformation or resampling can be a strong comparator when the desired variation is simple. The contribution of a learned editor should be a newly supported, validated class of variation or a better error–cost trade-off, not the relabeling of existing anatomy manipulation. This distinction keeps controlled sensitivity studies useful without overstating biological causality.

Figure 5 places 20 representative editing publications in the common 2018–2026* window; the arrows resolve to the bibliography [14, 48, 71–88]. The common selection and date conventions are defined in Section 1.

03 Editing

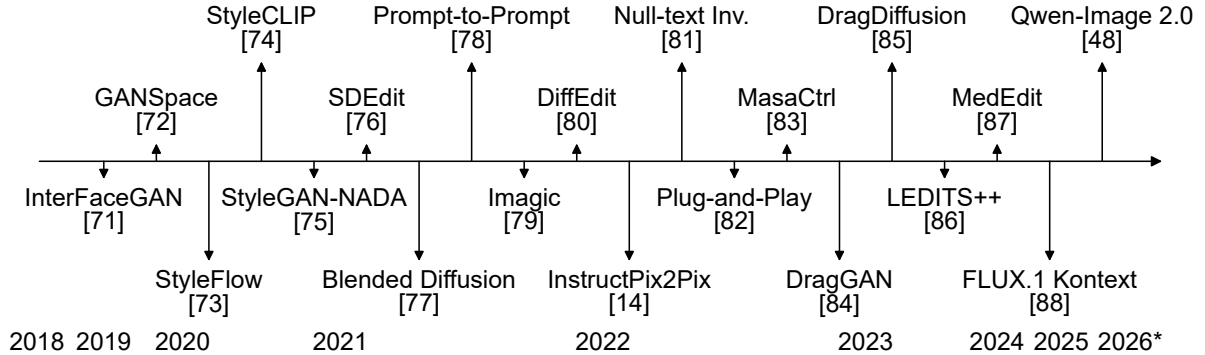


Figure 5: **03 Editing: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.1.4 04 Restoration / fill

Restoration retains an observed scene while compensating for degradation. Denoising, deblurring, colorization and filling missing regions differ in how much of the answer is identifiable. Palette uses a shared image-conditioned diffusion formulation across several such tasks [16]; multiple conditions do not alone establish multitask learning without shared task training and evaluation. DDRM reuses an image prior across linear inverse problems through operator structure [17]. Medical examples include microscopy restoration and MAISI tumor filling [9, 18]. A filled region may be a plausible simulation object rather than evidence of the true hidden anatomy. The task description must say whether the aim is retrospective recovery, prospective diagnosis, synthetic augmentation or experimental perturbation, because identical-looking outputs can have very different acceptable failure modes.

The unobserved region is where a strong prior can become a liability. A small residual on visible pixels says little about missing structures; several plausible samples do not automatically quantify calibrated uncertainty. Existing work already separates prior-induced hallucinations from ordinary reconstruction error and develops distribution-referenced hallucination metrics [18, 89]. The latter depends on a defined reference and approximation assumptions, not a universal truth oracle. Evaluation should therefore distinguish measurement-supported detail, ambiguous completion and wrong but convincing structures. For a clinical use, the question is whether the ambiguity changes a diagnosis or quantitative endpoint and whether the system can decline to fill it. Deterministic restoration remains an essential comparator: this operation-based taxonomy includes such comparators without claiming that every denoiser is a stochastic generative model.

Figure 6 places 20 representative restoration / fill publications in the common 2018–2026* window; the arrows resolve to the bibliography [16, 90–108]. The common selection and date conventions are defined in Section 1.

04 Restoration / fill

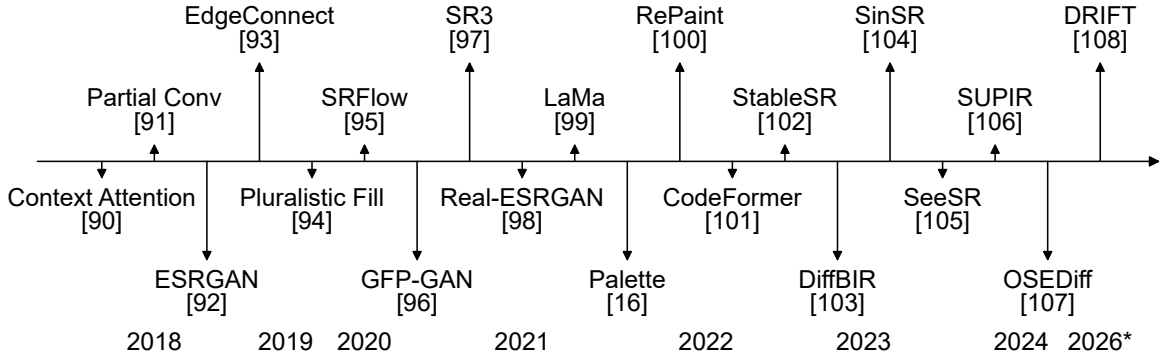


Figure 6: **04 Restoration / fill: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.1.5 05 Measurement inverse

Reconstruction uses a specified forward model and a particular measurement. CT projection data and MRI k-space constrain a latent object through Equation 2; they are not merely prompts for a similar-looking image. Generative compressed sensing makes representation error explicit, while DPS approximates a noisy likelihood using denoised estimates [5, 19]. Medical score-based reconstruction demonstrates reuse of an unconditional prior under different sampling operators [20]. In the inspected study, CT projections and single-coil MRI measurements are simulated from image data. Those experiments probe operator variation, but they do not constitute prospective scanner or diagnostic validation. A precise report separates real source images, simulated measurements, real acquisition tests, external patients and clinical readers rather than collapsing all of them into “real medical data.”

Faithfulness requires more than exact data fitting. Noise and model mismatch can make strict fitting undesirable, while a small data residual cannot certify components poorly constrained by the operator [89]. Generator-range bias can persist even with more observations [5], and posterior-sample diversity can mix model, numerical and observation effects. A stronger experiment fixes the raw-data split, compares a competitive non-generative method and a generative baseline, and measures a prespecified lesion, quantitative or reader endpoint at a realistic compute budget. Active acquisition is already explored with diffusion posterior samples [109]; novelty must lie beyond merely choosing the next view. Uncertainty becomes practically useful only when it supports a calibrated decision, rejection or additional measurement with a favorable total burden, including failed cases and reacquisition time.

Figure 7 places 20 representative measurement inverse publications in the common 2018–2026* window; the arrows resolve to the bibliography [17, 19, 20, 110–126]. The common selection and date conventions are defined in Section 1.

05 Measurement inverse

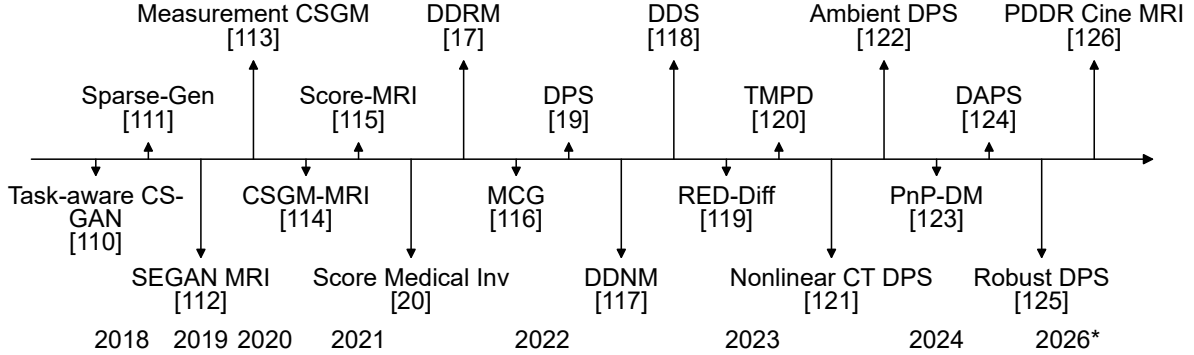


Figure 7: **05 Measurement inverse: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.1.6 06 Temporal generation

Time introduces several distinct problems. Unconditional video creates a sequence; interpolation fills missing times; prediction conditions on a past; action-conditioned generation also receives an intervention on the acquisition or environment. Video diffusion and Genie exemplify joint visual sequences and interactive dynamics [21, 22]; 4D-fy adds a dynamic 3D representation with complementary image and video supervision [23]. Medical cases include CHeart’s within-cycle anatomy, Cardiac Mesh Flow’s dynamic meshes and ultrasound views under probe movement [24–26]. A cardiac cycle is not longitudinal disease progression, and an autoregressive “causal” architecture is not a treatment-effect model. Similarly, a smooth rendered mesh is not automatically a measured displacement field. Each temporal use must specify its conditioning history, horizon, action semantics and independent reference.

Temporal validity needs negative controls that defeat superficial metrics. Cardiac Mesh Flow’s ablation shows that static hearts can attain strong periodic consistency, making endpoint agreement an insufficient motion validator [25]. CHeart’s dense temporal anatomy relies on processed segmentation labels, not independent local motion truth [24]. SonoWorld already grounds ultrasound generation in metric poses, while EchoWorld studies predictive representations for probe guidance [26, 127]; neither inspected evaluation establishes an independent real-robot clinical policy-transfer trial. Contact force, deformation, physiology and scanner settings can all break a pose-only observation model. Tests should include static outputs, condition shuffling, non-replayed actions and local motion or control endpoints. The important advance is not longer or smoother video, but a demonstrably correct relation between time, action and the downstream quantity.

Figure 8 places 20 representative temporal generation publications in the common 2018–2026* window; the arrows resolve to the bibliography [21, 128–146]. The common selection and date conventions are defined in Section 1.

06 Temporal generation

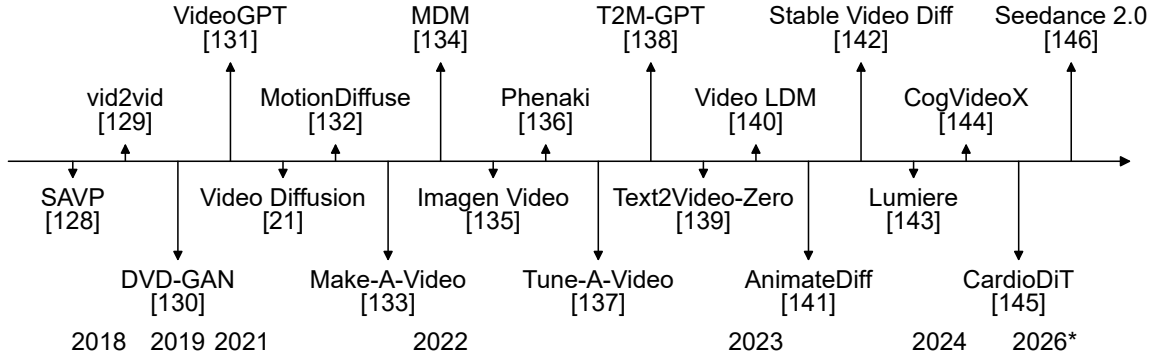


Figure 8: **06 Temporal generation: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.2 Conditioning and control

Conditions specify supplied information; control asks whether a requested property changes reliably. Text is one condition, not a synonym for conditioning. A model can use masks or words while still failing to preserve unrequested content. These distinctions explain the three rows below.

Figure 9 gives the examples for this coordinate of Figure 1.

02 Conditioning and control

Figure 1 coordinate 2 | Subitems 07, 08, 09

Machine learning

07 Conditional

Condition + noise -> samples
Select a conditional distribution
Keep given context, vary detail
Test adherence and diversity

08 Text-conditioned

Text -> new visual samples
Words specify desired content
Keep named attributes / relations
Test response to prompt edits

09 Controllable

Chosen factor -> steered sample
Change one requested property
Fix other inputs where possible
Test control and unwanted drift

Computer vision

— SPADE layout to scene

Street layout -> scene image
Condition on semantic regions
Keep layout; vary appearance
Test layout fit and realism

— Imagen text to scene

Text -> generated scene
Example: red cubes, blue ball
Target count, color, position
Test object / relation accuracy

— ControlNet spatial steering

Pose / edges / depth -> image
Guide the generated structure
Vary appearance, retain control
Test alignment / off-target drift

Medical image analysis

— MAISI anatomy to CT

Organ masks + spacing -> CT
Generate a synthetic 3D volume
Target requested anatomy
Test masks and CT plausibility

— RoentGen text to X-ray

Radiology prompt -> radiograph
Generate, not retrieve, an image
Target findings and anatomy
Test independent finding review

— MAISI anatomical steering

Selected organ mask -> CT
Anatomy control, not causation
Target the requested mask shape
Test changed / other structures

Figure 9: **Conditioning and control: ML–CV–medical correspondences.** Subitems 07, 08, 09. Meanings and illustrative uses; Need/Test labels are requirements, not evidence that every example passed. Sources: [9, 41, 147–149].

3.2.1 07 Conditional

Conditional. Conditional generation asks how outputs change when information is supplied, not how much training data are available. In ML, a condition selects a distribution: the same class or context can still admit many valid samples. SPADE transfers this idea to CV by synthesizing scenes from semantic layouts. MAISI provides a medical counterpart, generating synthetic three-dimensional CT with organ segmentations as conditions and configurable voxel spacing. The desired invariant is agreement with the supplied anatomy; unsupplied appearance remains variable. Text conditioning is one special case, while controllability additionally asks whether specific changes can be steered reliably. Suggested checks are condition agreement, diversity and anatomical plausibility, not assumed guarantees. A conditioning input does not establish patient identity, pathology truth, privacy or clinical benefit. The timeline and examples draw on the following sources [7, 9, 16, 30, 37, 39, 41, 52, 97, 147–157].

07 Conditional

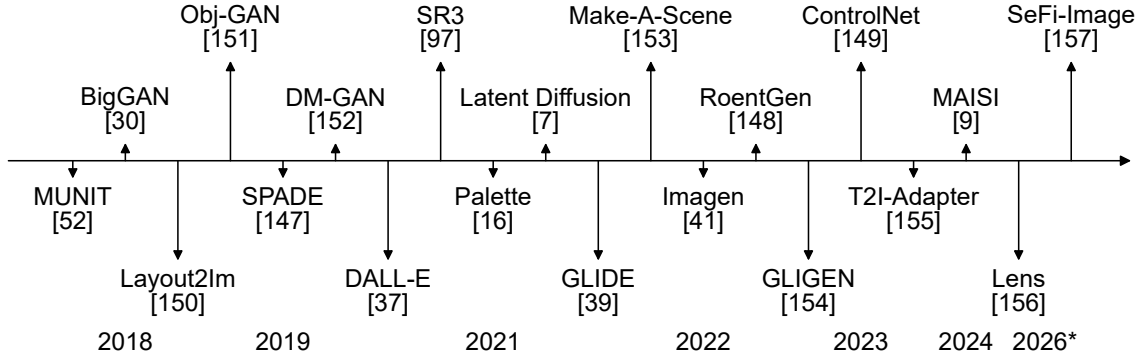


Figure 10: **07 Conditional: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.2.2 08 Text-conditioned

Text-conditioned. Text-conditioned generation uses language as the condition; it is therefore compatible with other conditioning modalities and data regimes. In ML, words specify requested attributes and relations, but the generator must translate them into visual content. Imagen provides a CV anchor for text-to-image diffusion, making object counts, colors and relations useful teaching probes rather than guaranteed successes. RoentGen adapts a pretrained latent diffusion model using chest radiographs and reports, then generates radiographs from medical-language prompts. Suggested checks distinguish prompt adherence from image realism and independently assess depicted findings and anatomy. A requested diagnosis is neither evidence that the output depicts it correctly nor truth about a patient. These abstract-supported examples establish generation capabilities, not complete pathology control, patient preservation or clinical proof. The timeline and examples draw on the following sources [7, 37, 39–42, 46, 47, 148, 151–153, 156–163].

08 Text-conditioned

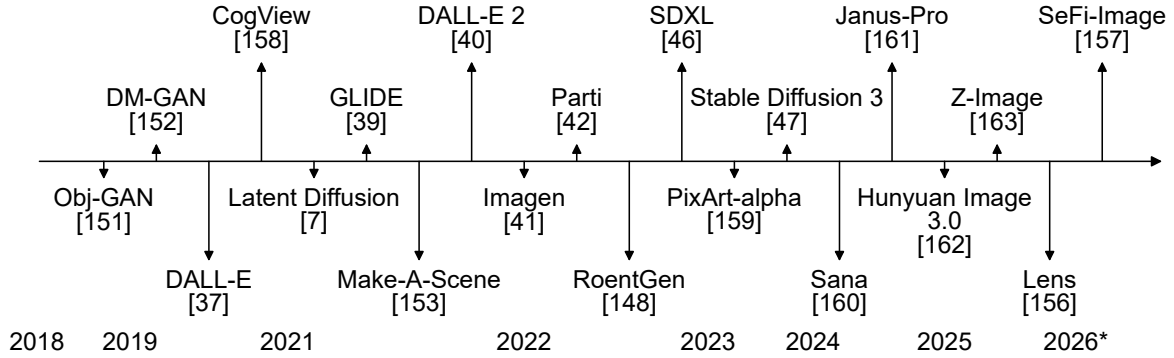


Figure 11: **08 Text-conditioned: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.2.3 09 Controllable

Controllable. Controllability is a steering objective, whereas conditioning describes information supplied to a generator. A useful ML probe varies one requested factor while fixing other inputs and, when meaningful, the random seed; outputs must still be checked for coupled changes. ControlNet adds trained spatial controllers to a pretrained text-to-image diffusion model using pose, edge or depth conditions. MAISI documents organ-mask conditioning for CT synthesis, providing a narrower anatomical-control example rather than a disease-intervention model. Suggested checks quantify response to the changed mask and unintended changes elsewhere. Neither a frozen base generator nor an anatomical control implies training-free control learning, causal identification, preserved patient identity or clinical validity. These coordinates can overlap with text conditioning and with either low-data adaptation or extensive training. The timeline and examples draw on the following sources [9, 14, 71–74, 78, 83, 147, 149, 153–155, 164–170].

09 Controllable

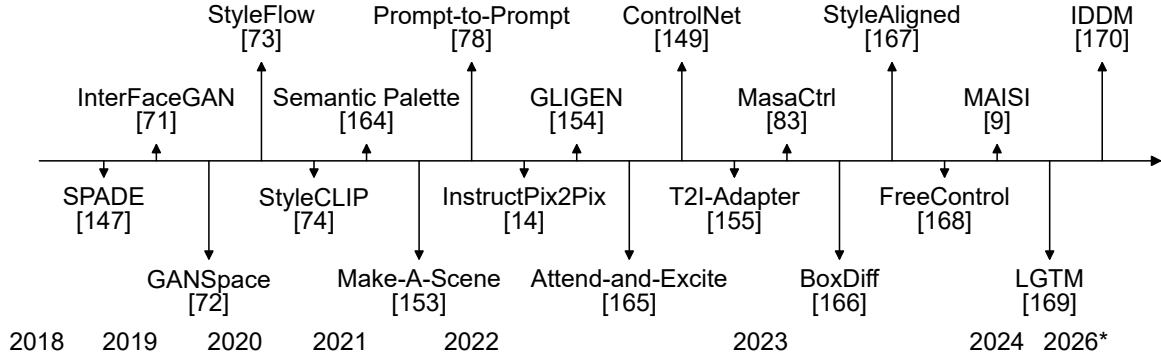


Figure 12: **09 Controllable: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.3 Data and adaptation regime

The budget must identify target examples, source pretraining and inference-time fitting separately. Few-shot limits a target support set; zero-shot or training-free reuse removes a specified fitting stage; data efficiency compares performance against a real-data budget. None alone establishes low total computation.

Figure 13 gives the examples for this coordinate of Figure 1.

03 Data and adaptation regime

Figure 1 coordinate 3 | Subitems 10, 11, 12

Machine learning

10 Few-shot

Few examples -> adapted model
Reuse a learned source prior
Keep target traits, vary samples
Test novelty and memorization

11 Zero-shot / training-free

New condition -> prior reuse
No target generator refitting
Pretraining can still be large
Specify what receives no training

12 Data-efficient

Limited real data -> generator
Reduce the real-example budget
Keep variation, avoid overfitting
Test quality versus data count

Computer vision

— DreamBooth personalization

Few photos -> new subject views
Fine-tune a pretrained generator
Target identity, not copied views
Test identity / novelty / prompts

— RePaint without retraining

Image + mask -> filled image
Reuse an unconditional DDPM
Modify sampling, not the network
Test visible pixels / filled region

— ADA limited-data GAN

Limited images -> trained GAN
Adapt critic-side augmentation
Compare at equal data budgets
Test diversity and sample quality

Medical image analysis

— EndoFSA endoscopy

Few lesion images -> adaptation
Reuse normal-image pretraining
Generate abnormal endoscopy
Test lesions and memorization

— Score-MRI sampling reuse

Sparse MRI data -> image
Reuse a magnitude-MRI prior
No sampling-pattern refitting
Test data fit and missed lesions

— EndoFSA target-data reuse

Scarce lesions -> adapted model
Reuse a normal-data prior
Keep pretraining budget explicit
Test real-patient utility / budget

Figure 13: **Data and adaptation regime: ML–CV–medical correspondences.** Subitems 10, 11, 12. Meanings and illustrative uses; Need/Test labels are requirements, not evidence that every example passed. Sources: [100, 115, 171–173].

3.3.1 10 Few-shot

Few-shot. Few-shot generation limits examples of a specified target, such as a subject or domain, rather than all data ever used. In ML, adaptation transfers a learned prior to that small support set; other methods can condition on support without weight updates. DreamBooth illustrates CV personalization by fine-tuning a pretrained text-to-image model on a few subject photographs. EndoFSA documents a medical case: a generator pretrained on abundant normal capsule-endoscopy images is adapted to scarce abnormal domains through restricted parameter updates. These are generated domain examples, not reconstructions of a particular patient’s lesion. Suggested checks include unseen-patient lesion review, diversity and memorization auditing. Exact medical shot counts and independent clinical performance require further evidence; few trainable parameters alone would not establish few-shot learning. The timeline and examples draw on the following sources [6, 170–172, 174–189].

10 Few-shot

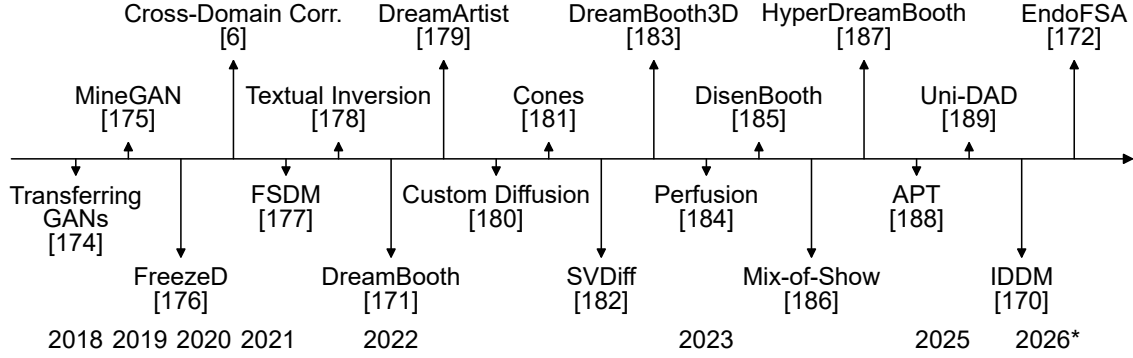


Figure 14: **10 Few-shot: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.3.2 11 Zero-shot / training-free

Zero-shot / training-free. This coordinate concerns reuse without fitting generator weights to the target condition or acquisition operator. It differs from semantic zero-shot recognition of unseen categories. No target labels also does not mean no pretraining or no target-modality images. In ML, an existing prior is combined with a new condition during inference. RePaint turns a trained unconditional diffusion model into an inpainter by changing reverse-sampling iterations rather than retraining its network. In medicine, Score-MRI combines a magnitude-image-trained score prior with data-consistency steps across sampling patterns. Its training-free aspect is acquisition-specific reuse, not learning MRI anatomy from nothing. Suggested tests assess measurement residuals and independently checked anatomy. Inference projections or latent optimization still incur computation; agreement with measured data cannot verify unmeasured lesions. The timeline and examples draw on the following sources [17, 19, 64, 76, 78, 80, 82, 83, 100, 115, 117, 139, 165–169, 190–192].

11 Zero-shot / training-free

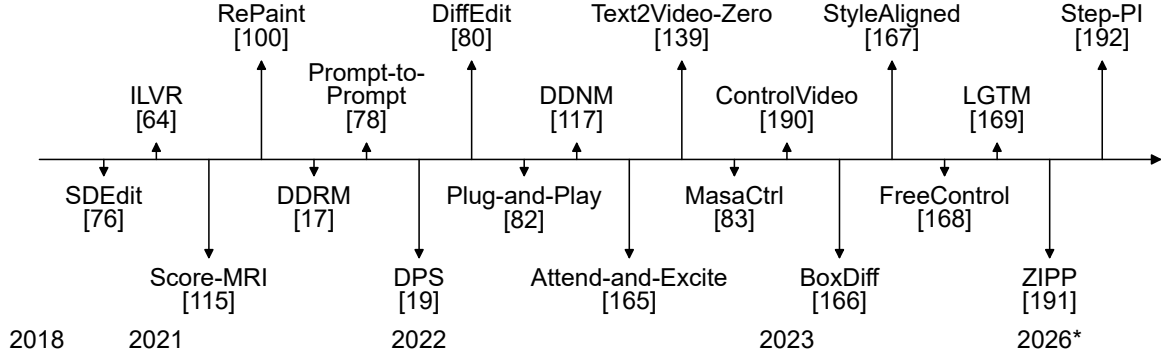


Figure 15: **11 Zero-shot / training-free: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate -1 , $+1$, -2 , $+2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.3.3 12 Data-efficient

Data-efficient. Data efficiency asks whether useful generation can be learned with fewer real examples or less target supervision at a specified quality level. In ML, regularization or transfer can reduce overfitting, but the data budget must be stated. ADA provides a CV anchor through adaptive discriminator augmentation for limited-data GAN training, including training from scratch and transfer. EndoFSA illustrates medical target-data reuse: normal-image pretraining supports adaptation to scarce abnormal endoscopy images. This is a documented low-data design, not a verified cross-method efficiency ranking. Suggested tests plot quality, diversity and independent downstream performance against real patient counts with fixed splits and comparable pretraining access. Count abundant pretraining separately; faster sampling, smaller adapters and more synthetic images do not themselves demonstrate lower real-data requirements. The timeline and examples draw on the following sources [6, 171–178, 180, 193–202].

12 Data-efficient

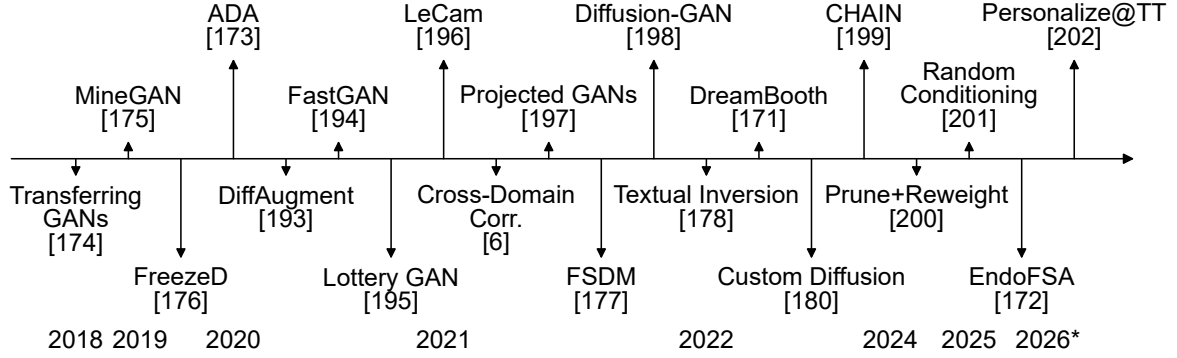


Figure 16: **12 Data-efficient: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.4 Task organization

Organization describes what shares learning and what belongs to the same sample. Multitask concerns objectives, multimodal concerns representations, and joint generation concerns dependent outputs. The same system can have all three properties; architecture or channel count alone establishes none of them.

Figure 17 gives the examples for this coordinate of Figure 1.

04 Task organization

Figure 1 coordinate 4 | Subitems 13, 14, 15

Machine learning

13 Multitask

Several tasks -> shared model
Share learning across objectives
Keep separate task evaluations
Test each task versus specialists

14 Multimodal

Different modalities -> modeling
Link language and visual content
Conditioning is not joint sampling
Test cross-modal agreement

15 Joint generation

Noise -> corresponding outputs
Learn how outputs depend
Describe one shared sample
Test pairs and individual outputs

Computer vision

— Palette generalist model

Task + input -> restored image
Colorize, inpaint, uncrop, restore
Share conditional diffusion
Test each task, not their average

— UniDiffuser image and text

Text -> image; image -> text
Also supports paired sampling
Specify the sampling mode
Test alignment in each mode

— JointNet RGB and depth

Text + noise -> RGB-depth pair
Couple two diffusion branches
Target matching scene geometry
Test edges and pair consistency

Medical image analysis

— AortaDiff shared objectives

Plain CT -> synthetic CT / masks
Synthesis + segmentation tasks
Target shared anatomical identity
Test image / Dice / size error

— UniX reports and X-rays

Report -> X-ray; X-ray -> report
Generate across distinct formats
Target findings, not patient ID
Test omissions / contradictions

— MedSegFactory image / mask

Text + noise -> image-mask pair
Cross-condition two streams
Target matching anatomy / labels
Test boundaries / real-task utility

Figure 17: **Task organization: ML–CV–medical correspondences.** Subitems 13, 14, 15. Meanings and illustrative uses; Need/Test labels are requirements, not evidence that every example passed. Sources: [16, 49, 203–206].

3.4.1 13 Multitask

Multitask. Multitask describes how objectives share learning, not how many channels enter a network. A generic ML example shares parameters across conditional generation objectives while retaining task-specific evaluation under shared or task-specific losses. Palette provides a CV instance: one generalist diffusion model handles colorization, inpainting, uncropping, and JPEG restoration. In medicine, AortaDiff combines synthetic contrast-enhanced CT generation from noncontrast CT with lumen and thrombus segmentation. It therefore also overlaps joint generation; coordinate values are not exclusive categories. Both image quality and segmentation or measurement error require separate checks against specialist baselines. The intended invariant is anatomical identity across tasks, not identical pixel values. Shared architecture alone does not establish a single unchanged checkpoint, and benchmark performance does not prove clinical substitution. The timeline and examples draw on the following sources [7, 16, 39, 49, 76, 203, 204, 207–220].

13 Multitask

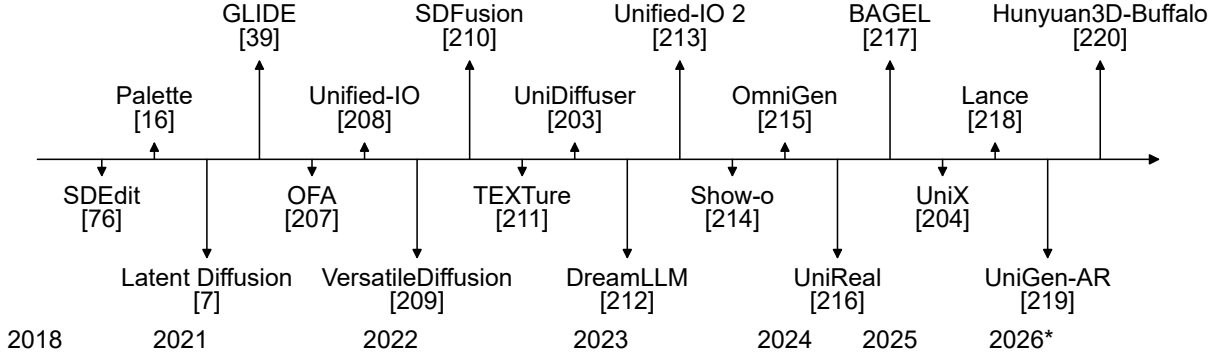


Figure 18: **13 Multitask: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.4.2 14 Multimodal

Multimodal. Multimodal generation connects semantically different representations, such as language and images. In ML, generating an image from text models a cross-modal conditional distribution; the text may remain fixed while only the image is sampled. UniDiffuser illustrates the CV distinction because the same framework supports image-to-text, text-to-image, and paired image-text generation through different sampling settings. UniX supplies a medical example with chest-radiograph report generation and text-guided radiograph synthesis. The invariant to test is agreement about the requested findings, not pixelwise identity with an individual patient. Multimodal conditioning alone is not joint sampling, and several MRI contrasts do not by themselves establish an any-to-any system. Required checks include finding-level omissions, contradictions, and unsupported details; a realistic image or fluent report is not clinical proof. The timeline and examples draw on the following sources [37, 39, 41, 203, 204, 212–214, 217, 218, 221–230].

14 Multimodal

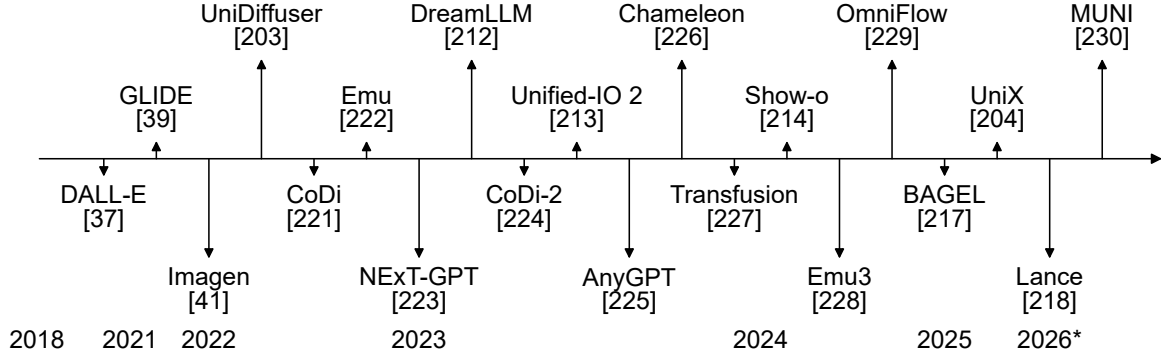


Figure 19: **14 Multimodal: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.4.3 15 Joint generation

Joint generation. Joint generation concerns dependence between outputs belonging to one sample. In ML, learning two plausible marginal distributions is insufficient: an image and its annotation must agree when sampled together. JointNet illustrates this in CV by coupling RGB and dense-modality diffusion branches to generate corresponding RGB-depth outputs. MedSegFactory provides a medical counterpart with two streams for images and segmentation masks, connected by joint cross-attention during denoising. Coupling can also use a shared latent or a dependent pipeline; simultaneous execution is not the defining requirement. Required checks assess each output and their correspondence, including boundaries, label semantics, and downstream performance on independent real cases. A jointly generated mask is not independent ground truth: both streams may reproduce the same anatomical error, and agreement alone cannot establish clinical validity. The timeline and examples draw on the following sources [203, 205, 206, 221, 229–244].

15 Joint generation

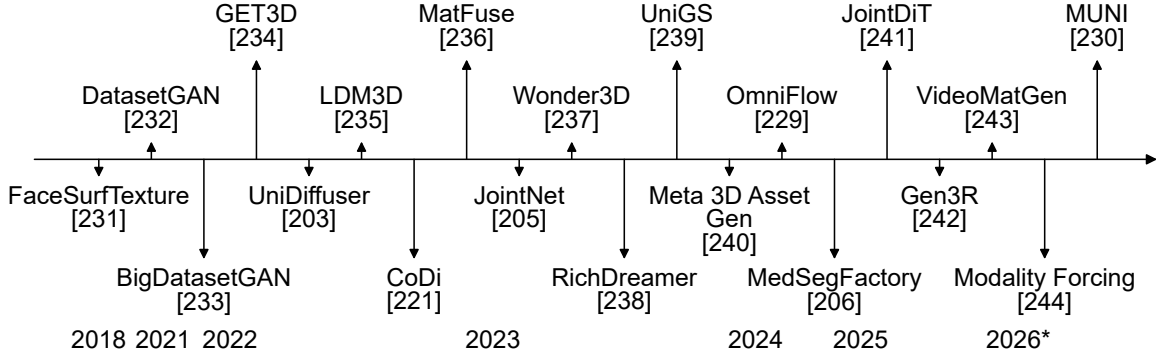


Figure 20: **15 Joint generation: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.5 Generated object

Objects specify what the output represents: an intensity array, appearance on a surface or domain, or geometry. A 3D CT volume remains an image, while a segmentation encodes anatomy. Tissue appearance is a medical texture analogue, not evidence of faithful surface texturing.

Figure 21 gives the examples for this coordinate of Figure 1.

05 Generated object

Figure 1 coordinate 5 | Subitems 16, 17, 18

Machine learning

16 Image

Noise -> sampled intensity array
Represent image appearance
Array meaning matters, not size
Test coverage / condition fit

17 Texture

Noise / example -> texture
Model appearance statistics
Geometry is fixed in this example
Test correlation and repetitions

18 Shape

Latent -> sampled geometry
Points / mesh / occupancy / SDF
Specify coordinates and topology
Test validity and coverage

Computer vision

— StyleGAN face images

Latent + noise -> RGB face
Vary pose and visual appearance
Raster output, not explicit mesh
Test diversity and distribution

— Text2Tex surface texture

Mesh + text -> texture map
Paint from multiple viewpoints
Keep supplied mesh geometry
Test seams / view consistency

— PointFlow point clouds

Latent -> new 3D shape
Sample points within each shape
No mesh connectivity specified
Test surface coverage / diversity

Medical image analysis

— MAISI synthetic CT

Masks + noise -> 3D CT image
Vary spacing and image size
Target anatomy and CT intensity
Test organ fit and intensity bias

— Virtual stain analogy

Autofluorescence -> stain image
GAN changes tissue appearance
Cell layout, not a mesh UV map
Test matched stained structures

— CAN-FLOW cardiac anatomy

Age / sex / BMI -> heart shape
Sample biventricular anatomies
Target metadata-related variation
Test phenotypes / shape validity

Figure 21: **Generated object: ML–CV–medical correspondences.** Subitems 16, 17, 18. Meanings and illustrative uses; Need/Test labels are requirements, not evidence that every example passed. Sources: [8, 9, 13, 28, 31, 245, 246].

3.5.1 16 Image

Image. Image specifies the generated object, not the learning mechanism or task. In ML, a model samples an array whose values represent intensity or color; generation can be unconditional, conditioned, or part of restoration. StyleGAN supplies a CV example by synthesizing face images with controllable appearance variation. MAISI demonstrates the medical target by using latent diffusion to synthesize three-dimensional CT, with organ segmentations available as ControlNet conditions. An intensity volume remains an image even though it occupies three-dimensional space; a segmentation volume instead encodes anatomical labels or shape. Required checks include anatomical agreement, intensity bias, diversity, and memorization rather than visual realism alone. An organ mask is a control signal, not clinical evidence, and a synthetic CT must not be treated as an acquired patient measurement. The timeline and examples draw on the following sources [7, 9, 30–34, 36–41, 44, 46, 47, 159–161, 219, 247].

16 Image

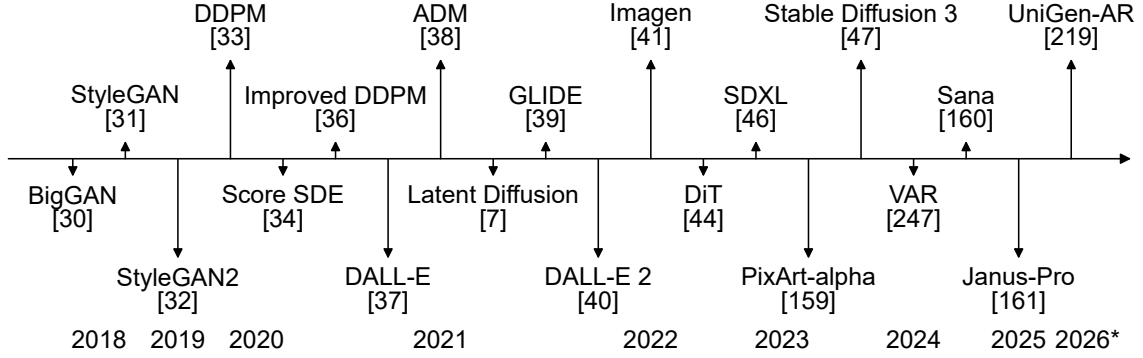


Figure 22: **16 Image: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.5.2 17 Texture

Texture. Texture names generated appearance, which may be a spatial pattern, a mesh color map, or a material field. The ML problem models spatial appearance statistics rather than geometry alone. Text2Tex is a direct CV example: a supplied mesh and text guide multiview diffusion inpainting to build its texture. Mesh geometry is the fixed support; seams and cross-view agreement need checking. The medical bridge here is explicitly an appearance analogy: virtual staining maps label-free tissue autofluorescence to a stained-looking microscopy image using a GAN. It generates image appearance, not a UV texture for a mesh. Likewise, MIVAE couples MRI appearance with cardiac shape; that does not make MRI intensity a surface material. Matching histological structures and disease-relevant details requires independent validation beyond visual realism. The timeline and examples draw on the following sources [8, 13, 211, 231, 234, 236, 243, 245, 248–261].

17 Texture

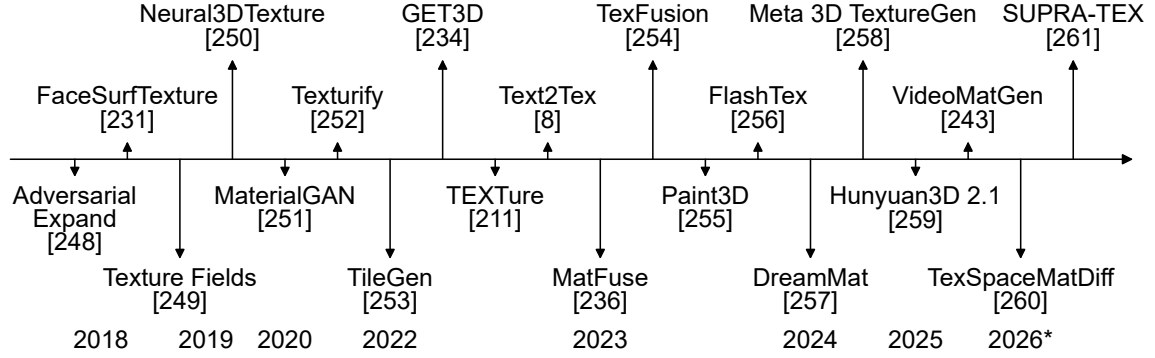


Figure 23: **17 Texture: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.5.3 18 Shape

Shape. Shape specifies a geometry-bearing output, not a particular generator. In ML, learning a distribution of coordinates, occupancies, or signed distances permits new spatial configurations to be sampled. PointFlow illustrates the CV case by modeling a distribution of shapes and a conditional distribution of points for each shape. Its point clouds do not themselves provide triangle connectivity. CAN-FLOW supplies a medical example: a conditional normalizing flow models metadata-dependent cardiac shape representations and generates biventricular anatomies. It is not a flow-matching method merely because its name contains flow. Required checks cover anatomical validity, phenotype distributions, subgroup variation, and geometric representation conventions. Evidence from healthy training anatomies does not establish diseased anatomy coverage, physiological function, or simulation readiness. Appearance and temporal motion remain separate, potentially overlapping coordinates. The timeline and examples draw on the following sources [28, 210, 220, 234, 246, 262–277].

18 Shape

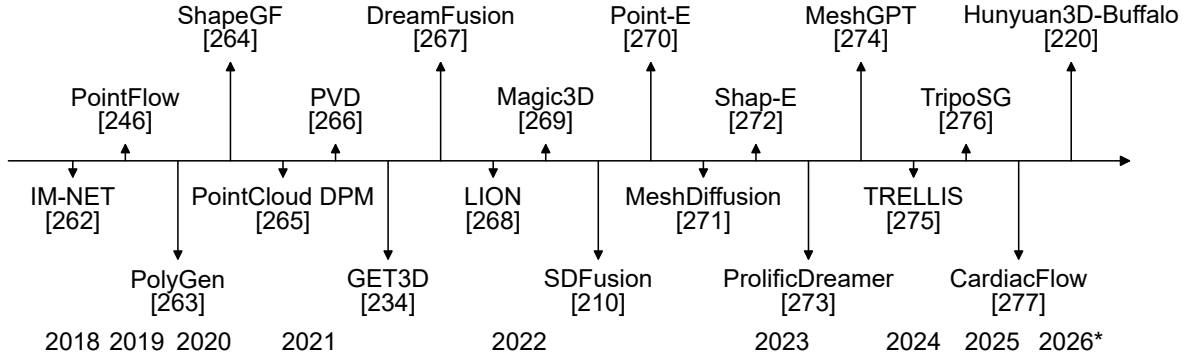


Figure 24: **18 Shape: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.6 Temporal structure

Temporal structure asks how states depend on time and actions. A time-indexed 3D object, a missing intermediate frame and a future observation under a chosen action require different references. Their smoothness can be similar while their physical or control meaning differs.

Figure 25 gives the examples for this coordinate of Figure 1.

06 Temporal structure

Figure 1 coordinate 6 | Subitems 19, 20, 21

Machine learning

19 4D / 3D+t

Latent -> 3D states over time
One 3D object changes in time
Keep space-time dependence
Test held-out temporal states

20 Frame interpolation

Endpoints + time -> interior state
Fill the observed time interval
Keep endpoints and time order
Test withheld intermediate states

21 Action / world model

History + action -> next state
Learn dynamics, not a policy
Retain response to given actions
Test new actions / rollout drift

Computer vision

— MAV3D dynamic scene

Text -> dynamic radiance field
Render at chosen time and view
Keep cross-view object identity
Test view and motion coherence

— FILM intermediate frames

Two images -> interior frames
Handle large in-between motion
Keep identity across endpoints
Test occlusion / withheld frames

— GameNGen interactive views

Past frames + actions -> frame
Generate an interactive rollout
Keep state under player input
Test action effects, not replay

Medical image analysis

— CHeart cardiac anatomy

Clinical factors -> 4D anatomy
Generate a cardiac sequence
Target anatomy / cycle structure
Test true volume-time curves

— DDM cardiac MR interpolation

Diastole / systole -> interim MR
Learn a deformation-based path
Target source anatomy / topology
Test acquired in-between phases

— SonoWorld probe control

History + pose -> ultrasound
Ground control in tracked scans
Target pose-dependent anatomy
Test new actions / round-trip drift

Figure 25: **Temporal structure: ML–CV–medical correspondences.** Subitems 19, 20, 21. Meanings and illustrative uses; Need/Test labels are requirements, not evidence that every example passed. Sources: [24, 26, 278–282].

3.6.1 19 4D / 3D+t

4D / 3D+t. This coordinate asks whether three spatial dimensions evolve over physical time, rather than whether a sampler runs for many numerical steps. At the ML level, a model generates dependent spatial states indexed by time; independent volumes do not establish a coherent trajectory. In CV, MAV3D converts text into a dynamic 3D radiance field that can be rendered from different viewpoints. In medicine, CHeart conditions generated cardiac anatomical sequences on clinical factors. These examples illustrate an output structure, not an exclusive model family, and can also carry conditioning or mechanism labels. Validation should compare anatomy and motion against held-out reference sequences, including ventricular volume trajectories where appropriate. Smoothness or plausible population variation does not establish measured patient physiology, electrical activation, or a valid individual digital twin. The timeline and examples draw on the following sources [23, 24, 145, 278, 280, 283–297].

19 4D / 3D+t

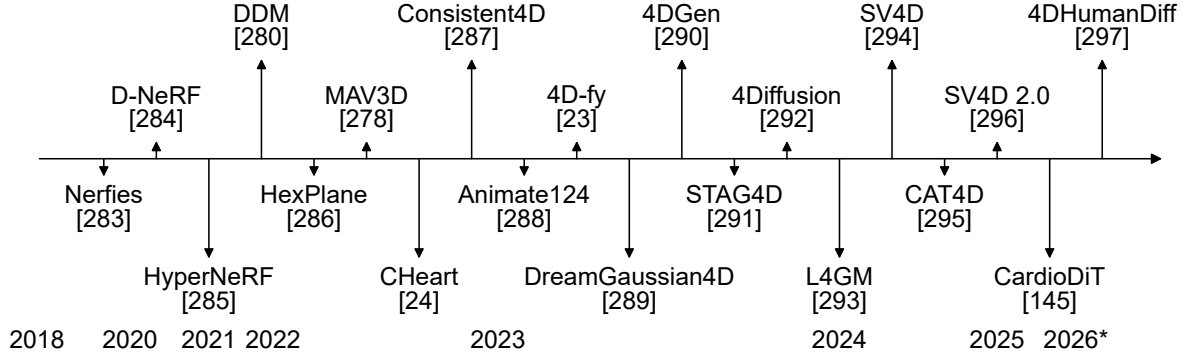


Figure 26: **19 4D / 3D+t: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.6.2 20 Frame interpolation

Frame interpolation. Frame interpolation fills times strictly between observed endpoints; extrapolation predicts beyond them. The ML problem is endpoint-conditioned synthesis, potentially with several plausible intermediate paths. FILM supplies the CV example by synthesizing intervening image frames even with large motion. The medical DDM example instead generates intermediate 3D cardiac MR volumes between diastolic and systolic observations using learned deformation. Thus the same temporal coordinate can describe ordinary image frames or spatial volumes, and it overlaps 4D generation when the latter evolve over time. Evaluation should withhold genuinely acquired intermediate phases and check anatomy, timing, and motion, rather than score endpoint reproduction alone. Interpolating a visually smooth heartbeat does not determine that patient’s actual intermediate physiology or justify predicting disease progression between clinical visits. The timeline and examples draw on the following sources [279, 280, 298–315].

20 Frame interpolation

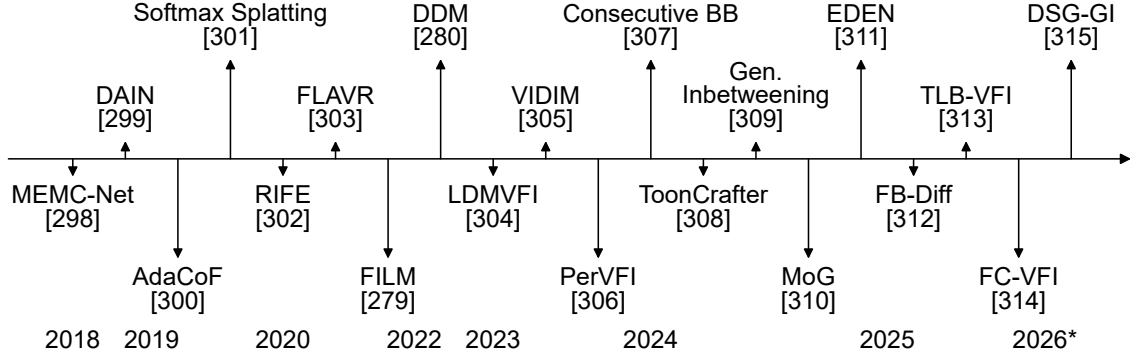


Figure 27: **20 Frame interpolation: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate -1 , $+1$, -2 , $+2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.6.3 21 Action / world model

Action / world model. An action-conditioned model predicts what observation or state follows a specified action and history; a policy instead chooses the action. World Models illustrates the ML formulation with a probability distribution over the next latent state conditioned on the current action and memory. GameNGen turns past frames and player actions into the next game frame. SonoWorld is a documented ultrasound preprint that grounds video generation in metric probe poses using EM-tracked freehand scans. Its pose-control evaluation is not proof that every generated response to a new action is physically valid. Validation should compare held-out action responses, opposite-direction controls, and accumulated rollout drift. Replaying recorded poses or producing convincing videos does not establish causal intervention validity, safe robotic control, or clinical benefit. The timeline and examples draw on the following sources [22, 26, 281, 282, 316–331].

21 Action / world model

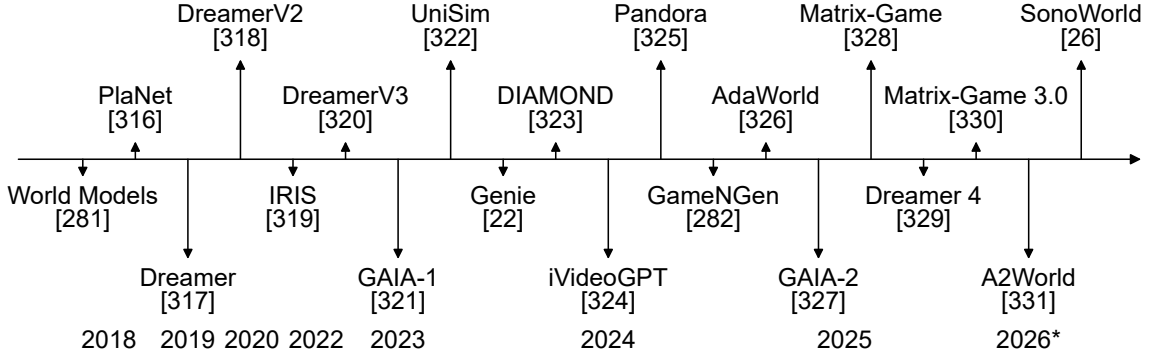


Figure 28: **21 Action / world model: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.7 Modeling mechanism

Mechanisms specify objectives and probability paths, not tasks or backbones. GAN, diffusion and flow matching are the three plotted anchors, not an exhaustive mechanism taxonomy. VAEs, autoregressive models and normalizing flows remain essential foundations [4, 33, 34, 332–335]. The following histories explain what changes mechanistically without implying that a newer family automatically preserves more information.

Figure 29 gives the examples for this coordinate of Figure 1.

07 Modeling mechanism

Figure 1 coordinate 7 | Subitems 22, 23, 24

Machine learning

22 GAN

Latent code -> synthetic sample
Generator competes with critic
Goal: match the data distribution
Test realism and coverage

23 Diffusion

Noise -> denoised sample
Learn reversal over noise levels
Goal: learn the data distribution
Test quality / coverage / privacy

24 Flow matching

Sample + path time -> velocity
Learn source-to-data transport
Match a chosen probability path
Test velocity fit / coverage

Computer vision

— StyleGAN image synthesis

Latent styles -> face image
Control attributes across scales
Target scale-specific appearance
Test realism / latent coverage

— Latent diffusion synthesis

Text + latent noise -> image
Denoise a compressed code
Target content after decoding
Test prompts and fine details

— Rectified-flow transformer

Text + noise -> generated image
Learn transport on chosen paths
Target requested visual content
Test prompts at matched cost

Medical image analysis

— MedGAN PET-to-CT

PET image -> synthetic CT
Adversarial and feature losses
Target source anatomy retention
Test attenuation error / review

— Medical Diffusion volumes

Noise -> 3D MRI or CT
Denoising diffusion synthesis
Check anatomy / slice coherence
Test pathology / memorization

— Medical flow matching

Noise + class / mask -> image
Documented 2D / 3D synthesis
Target conditional anatomy
Test masks and lesion fidelity

Figure 29: **Modeling mechanism: ML–CV–medical correspondences.** Subitems 22, 23, 24. Meanings and illustrative uses; Need/Test labels are requirements, not evidence that every example passed. Sources: [7, 31, 33, 34, 43, 47, 53, 332, 335, 336].

3.7.1 22 GAN

GAN. A GAN is defined by adversarial learning between a generator and discriminator, not by a specific network architecture or imaging task. The ML example maps random latent codes into samples while discriminator feedback supplies a learned training signal. StyleGAN applies this principle to image synthesis with scale-specific style control. MedGAN applies adversarial, feature, and style losses to PET-to-CT translation for attenuation correction rather than direct CT diagnosis. Here the source scan constrains the desired anatomical content, but those losses do not guarantee that lesions or quantitative attenuation values survive translation. Evaluation should measure distribution coverage as well as realism, then check paired anatomy, quantitative errors, and task-specific reader judgments. A discriminator score is neither diagnostic confidence nor evidence that all clinically important modes are represented. The timeline and examples draw on the following sources [30–32, 52, 53, 58, 92, 147, 173, 193, 198, 332, 337–345].

22 GAN

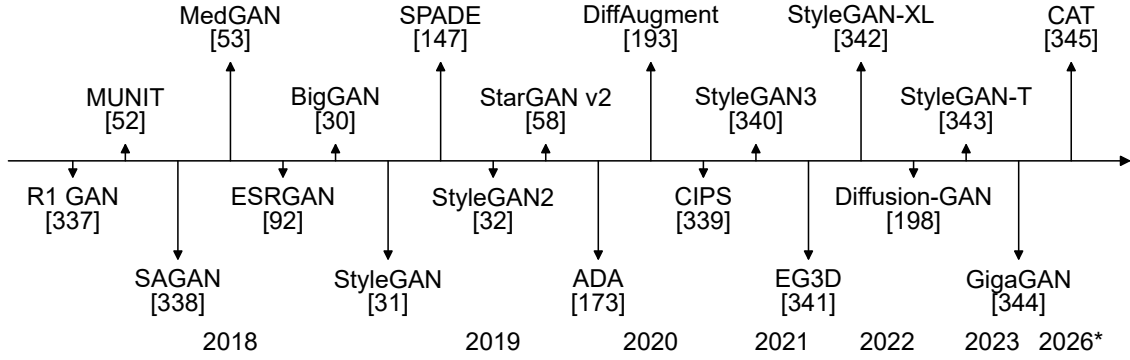


Figure 30: **22 GAN: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.7.2 23 Diffusion

Diffusion. Diffusion models learn how to reverse progressive corruption, commonly through denoising or noise-dependent score estimation. The ML example starts from noise and repeatedly applies the learned reverse process. Latent diffusion performs this process in a compressed representation and decodes it into an image, with text or other inputs providing optional conditions. Medical Diffusion documents three-dimensional MRI and CT synthesis. Its abstract reports reader assessment of appearance, anatomical correctness, and slice consistency; this is narrower than diagnostic equivalence or privacy protection. Validation should separately assess anatomical structure, rare abnormalities, memorization, and performance on independent clinical tasks. Numerical sampling time is not patient time. A deterministic probability-flow ODE can also sample a score-trained diffusion model, so using an ODE alone does not establish flow-matching training. The timeline and examples draw on the following sources [7, 19, 33, 34, 36, 38–41, 43, 44, 46, 97, 149, 346–351].

23 Diffusion

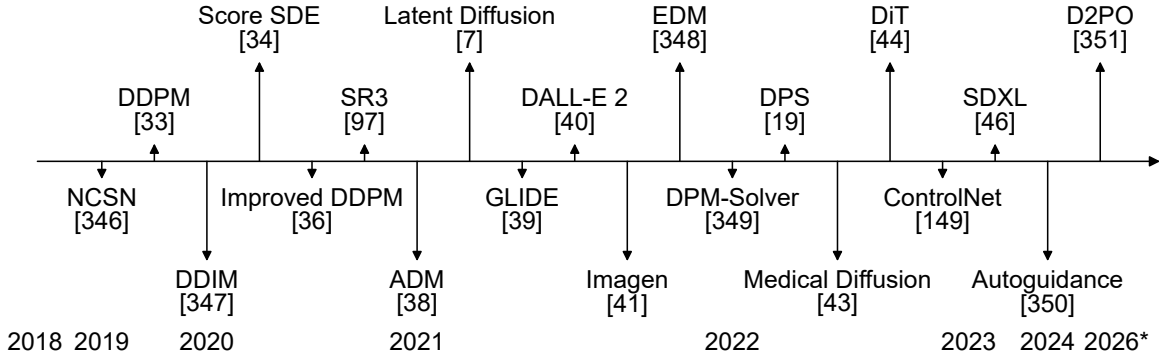


Figure 31: **23 Diffusion: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.7.3 24 Flow matching

Flow matching. Flow matching trains a velocity field against targets defined along chosen source-to-data probability paths. At the ML level, the input is an intermediate sample and path time, and the output is a transport direction used to generate new samples. Rectified-flow transformers provide the CV example through text-to-image synthesis. Flow Matching for Medical Image Synthesis documents an optimal-transport variant for 2D and 3D outputs with class or mask conditioning. This is a training coordinate, not a synonym for any invertible network, optical-flow estimator, or ODE solver. It can share probability paths and sampling machinery with diffusion. Validation should check transport accuracy, condition adherence, and anatomy at a declared sampling budget. The abstract-supported medical application does not establish clinical benefit, exact anatomical preservation, or universally faster generation. The timeline and examples draw on the following sources [47, 335, 336, 352–368].

24 Flow matching

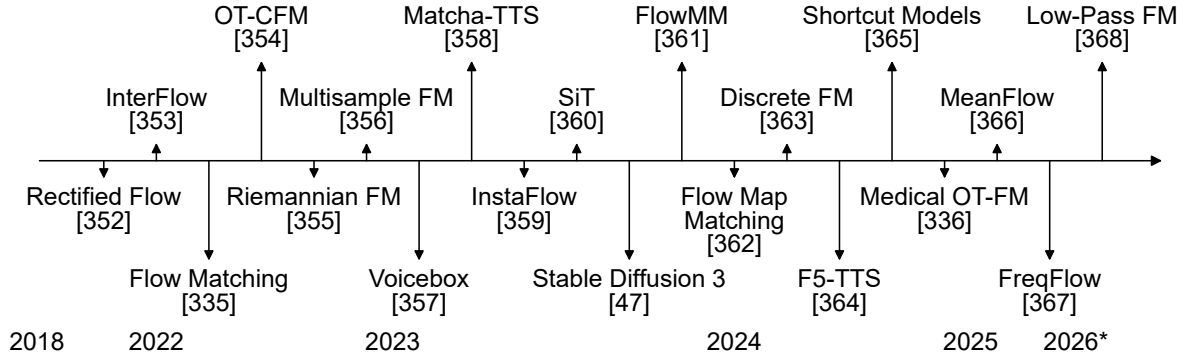


Figure 32: **24 Flow matching: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.8 Requirements and evaluation

Requirements concern what must remain correct and what resources the useful workflow consumes. Fidelity needs a reference, efficiency a budget, and data consistency a forward model with noise assumptions. These complementary tests cannot substitute for one another, and a favorable average may conceal consequential failures.

Figure 33 gives the examples for this coordinate of Figure 1.

08 Requirements and evaluation

Figure 1 coordinate 8 | Subitems 25, 26, 27

Machine learning

25 Fidelity

Real / synthetic data -> scores
PRDC: fidelity versus coverage
Declare the reference features
Test outliers / missing modes

26 Efficiency

Same workload -> cost / quality
Time / memory / training budget
Keep a declared quality target
Test full runtime, same hardware

27 Measurement consistency

Data + operator -> reconstruction
A prior fills unobserved parts
Respect the stated noise model
Test residual / unresolved detail

Computer vision

— GenEval compositional checks

Prompt + image -> alignment test
Check count, color and position
Target requested object relations
Test attribute binding, not looks

— DPM-Solver sampling

Trained diffusion prior -> image
Use fewer model evaluations
Check quality at each budget
Test wall time, not step count

— DDNM null-space restoration

Linear observation -> image
Generate null-space content
Keep noiseless measurements
Test data fit and structure

Medical image analysis

— Pathologist-guided alignment

Synthetic tissue -> expert review
Reward-guided diffusion tuning
Target meaningful tissue features
Test independent diagnostic truth

— MAISI-v2 volume synthesis

Noise + conditions -> 3D image
Rectified flow speeds synthesis
Check anatomy and condition fit
Test total time, failures, retries

— Score-based CT / MRI

Measured data -> image draws
Supply the operator at test time
Target noise-aware data fit
Test residuals / hidden anatomy

Figure 33: **Requirements and evaluation: ML–CV–medical correspondences.** Subitems 25, 26, 27. Meanings and illustrative uses; Need/Test labels are requirements, not evidence that every example passed. Sources: [19, 20, 27, 117, 349, 369–371].

3.8.1 25 Fidelity

Fidelity. Fidelity must name its reference: a data distribution, an input condition, a perceptual target, or a clinically meaningful feature. At the ML level, the density and coverage metrics associated with PRDC distinguish sample fidelity from distribution coverage rather than collapsing them into one score. GenEval supplies the CV example by checking generated object counts, colors, and spatial relationships against text prompts; it is an evaluator, not a generator. The medical human-feedback framework uses pathologist judgments, a learned reward model, and diffusion fine-tuning to improve clinical plausibility. A convincing texture can still conceal an incorrect cell type or missing abnormality. Evaluation should therefore include independent expert and task-specific checks under patient-separated testing. Expert preference or better generic image scores alone does not establish diagnostic fidelity or prospective utility. The timeline and examples draw on the following sources [38, 369–387].

25 Fidelity

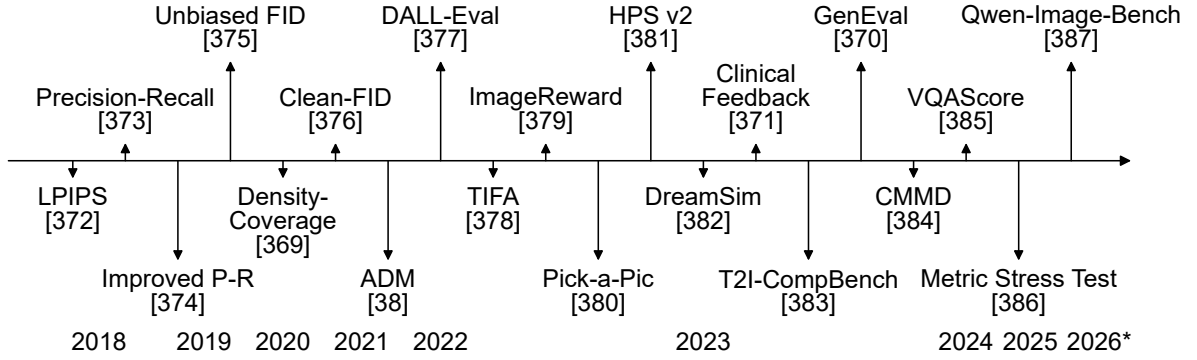


Figure 34: **25 Fidelity: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.8.2 26 Efficiency

Efficiency. Efficiency asks which resource is saved at what output quality, not simply how many sampling steps are removed. The ML comparison fixes data, hardware, and a quality target before reporting latency, memory, or training cost. DPM-Solver illustrates CV inference acceleration through a specialized solver for diffusion ODEs without additional training. MAISI-v2 documents accelerated 3D medical image synthesis using rectified flow and a region-specific contrastive loss; it is not merely DPM-Solver applied to medical scans. These mechanisms can occupy other coordinates at the same time. Evaluation should include encoding, decoding, loading, postprocessing, quality control, failures, and retries when measuring the total workflow. A fast sampling kernel or fewer network evaluations does not prove lower end-to-end cost, preserved lesion information, or diagnostic noninferiority. The timeline and examples draw on the following sources [7, 27, 36, 45, 347–349, 352, 366, 388–399].

26 Efficiency

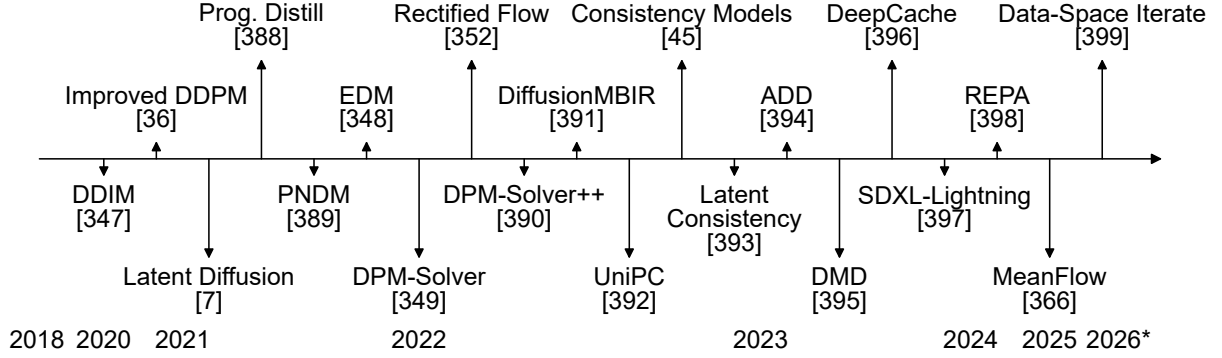


Figure 35: **26 Efficiency: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

3.8.3 27 Measurement consistency

Measurement consistency. Measurement consistency tests a reconstruction against acquired data through a specified forward operator and noise model. The ML example constrains observed components while a generative prior supplies unresolved ones. DDNM illustrates the CV case by modifying null-space content during diffusion-based linear image restoration. The medical score-based inverse framework reconstructs CT or MRI from measurements using a physical measurement model at test time. DPS additionally illustrates why noisy observations need a noise-aware objective rather than an indiscriminate exact-fit projection. Validation should examine residuals using the stated noise statistics and compare clinically relevant structures against independent references. A small residual cannot identify anatomy in an operator’s null space: different images may produce identical observations. Neither measurement fit nor a plausible posterior sample alone establishes unobserved truth or calibrated uncertainty. The timeline and examples draw on the following sources [17, 19, 20, 100, 114–117, 121, 122, 124, 391, 400–407].

27 Measurement consistency

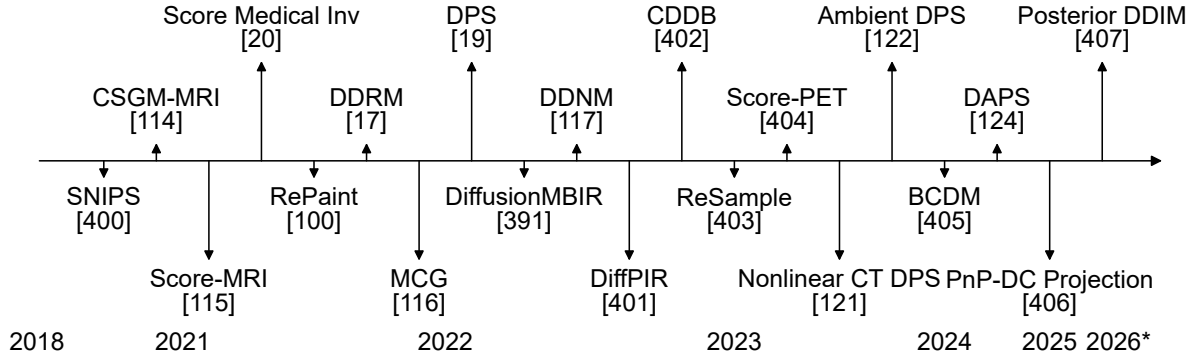


Figure 36: **27 Measurement consistency: representative publication timeline.** Twenty equally spaced papers in chronological order; year spacing is nonuniform. Black arrows alternate $-1, +1, -2, +2$. 2026* is YTD through 22 October. Reference numbers identify the publications; Section 1 defines selection and dates.

4 Cross-Dimensional Trends and Evidence Synthesis

4.1 What Changes, and What the Curves Cannot Establish

Mechanisms change faster than the underlying questions. Figure 1 shows declining GAN vocabulary and rising diffusion vocabulary before the current-year endpoint, but synthesis, reconstruction and control remain distinct operations under either mechanism. In complete-frame binary mentions, GAN falls from 176/428 (41.12%) in 2018 to 226/6,219 (3.63%) in 2025; diffusion rises from 0/428 to 2,353/6,219 (37.84%). Flow matching appears in 231/6,219 records (3.71%) in 2025. These shifts describe the query’s language, not a controlled comparison of accuracy, cost or clinical usefulness. A reader choosing a reconstruction problem should first identify its information and error constraints, then select a mechanism, rather than reverse that order.

2026 emphasizes action-related language, but is not a complete-year growth test. Among 7,019 YTD records, temporal generation matches 1,867 (26.60%), action conditioning 2,086 (29.72%), diffusion 1,544 (22.00%) and flow matching 401 (5.71%). Mean concept TF-IDF is 0.1803 for action conditioning versus 0.0739 in 2025, and 0.0600 for diffusion versus 0.1068. Query composition, indexing and writing choices can change these quantities. In particular, the action concept includes work outside temporal-generation matches, and the fixed 102-concept normalization couples panel magnitudes. A higher action curve does not show that models have learned useful interventions; that requires action-conditioned tests against replay and other strong controls.

Robust direction is more defensible than precise field share. Historical prevalence sensitivity places the 2018–2025 GAN change between -50.17 and -20.47 percentage points, diffusion between $+12.08$ and $+58.30$, and flow matching between $+1.51$ and $+5.71$. These are specification ranges, not confidence intervals. Direction survives these tested definitions, while magnitude does not. The 96-record relevance audit explains why interpreting a single filtered denominator as the literature would be misleading. Consequently, this survey uses counts to locate questions for evidence inspection, not to rank the scientific value of research topics.

4.2 Combinations Matter More Than Isolated Keywords

Cross-coordinate matches show which combinations deserve reading. We preselected five concept pairs spanning mechanism, condition, data regime, organization, operation and requirements, then re-evaluated the frozen rules over all 25,056 abstracts. All 918 annual concept counts reproduced exactly. Table 2 divides joint matches by anchor matches, so its denominator differs from the atlas and each row answers a specific conditional question. It is a diagnostic view of written claims, not evidence that the corresponding components were implemented together.

Table 2: **Cross-coordinate co-mentions in the fixed visual-generation frame.** Each entry is both-label matches divided by first-label matches in that year (percentage). The full yearly frames contain 6,219 records in 2025 and 7,019 in 2026*. *YTD retrieved 21 October. Labels are overlapping title/abstract matches, not verified method adoption; the years are not matched-calendar exposure.

Anchor / partner	2025	2026*
Diffusion / text-conditioned	1243/2353 (52.8%)	753/1544 (48.8%)
Temporal generation / action	139/1519 (9.2%)	369/1867 (19.8%)
Multimodal / text-conditioned	520/1089 (47.8%)	422/1040 (40.6%)
Few-shot / text-conditioned	35/77 (45.5%)	27/68 (39.7%)
Measurement inverse / data consistency	8/117 (6.8%)	10/123 (8.1%)

The action shift is not just more temporal papers. Action terms occur in 139/1,519 temporal-generation matches (9.2%) in 2025 and 369/1,867 (19.8%) in 2026 YTD. This supports inspecting the transition from producing sequences to predicting consequences of supplied actions. It does not identify which actions are measured, whether hidden state is observed, or whether a policy transfers. By contrast, text-conditioned terms occur in roughly half of diffusion matches in both snapshots (52.8% and 48.8%): diffusion and text-to-image should not be treated as synonyms. Similar care applies to multimodal and few-shot rows; their text co-mentions do not establish any-to-any generation or a comparable pretraining budget.

Sparse vocabulary is a search warning, not an empty field. Explicit data-consistency phrases co-occur in only 8/117 measurement-inverse matches in 2025 and 10/123 in 2026. This cannot support the claim that most reconstruction methods ignore measurements: likelihoods, forward operators, projections and domain-specific terms may express the same idea without these words. Likewise, small joint-generation or interpolation curves do not establish unstudied applications. A candidate gap must survive synonym expansion, direct-neighbor reading and a check of whether the proposed improvement is already expressed under another coordinate.

4.3 From Capability to Supported Use

There is no single scalar maturity of a subtask. Implementation, external image evaluation, clinical-reader agreement, workflow feasibility and patient outcomes answer different questions. Table 3 attaches these distinctions to named studies. It records what selected studies actually test, rather than assigning their evidence to all papers sharing a task label. A clinical non-diffusion MRI comparator can establish that a particular accelerated workflow is testable, but cannot validate a new diffusion sampler. Conversely, a recent generative method need not be clinically deployed to offer a useful, bounded technical result.

These selected sources span different operations and levels of validation. They are not a representative sample of all 27 subitems, and their limitations must remain attached to the named workflow rather than generalized to a whole mechanism.

Table 3: **Source-specific evidence ledger.** Scope: selected medical studies and clinical comparators, not a field-wide maturity ranking. Inspection locations and hashes are in the evidence log.

Source / use	What is evaluated	What does not follow
MAISI [9]	Conditional volumes; external autoPET image-distribution and downstream segmentation tests.	Diagnostic benefit, rare-pathology coverage or a privacy guarantee.
Medical score inversion [20]	CT/MRI inverse tasks with simulated projections or single-coil measurements.	Prospective scanner or independent reader validation.
Accelerated knee MRI [408]	Prospective 170-participant, six-reader interchangeability study.	Surgical truth, cross-vendor performance or validation of new diffusion methods.
MR-only brain RT [12]	Workflow completed in 20/21 cases with artifact-related fallback.	Universal synthetic-CT reliability or improved patient outcomes.
UV-PARS virtual histology [409]	Diagnostic concordance and blinded quality comparison by three pathologists.	Surgical-margin or patient-outcome benefit.
CAN-FLOW [28]	Metadata-conditioned static healthy anatomy; subgroup and tail comparisons.	Diseased dynamics or a device-trial endpoint.
Cardiac Mesh Flow [25]	Dynamic mesh benchmarks and static-heart periodicity control.	Independent local motion truth or patient-specific physiology.
VICTRE [410]	Procedural/physics virtual mammography comparison; 2,986 virtual patients.	General validity of learned cohorts, malignancy diagnosis or mortality prediction.
SonoWorld [26]	Pose-grounded ultrasound; 42 subjects, 33/9 split; action-selection/replay tests.	Independent real-robot clinical policy transfer or force-valid dynamics.

Three general priorities follow across ML and CV, before medicine adds further constraints. First, test compositional generalization: can independently useful conditions be combined without breaking an invariant? Second, compare total resources, including pretraining access, target examples and per-sample inference, rather than calling a frozen or small adapter universally efficient. Third, choose an evaluator that can disagree with the generator. These are synthesis priorities, not verified novel research gaps. JointNet already couples visual outputs, MedSegFactory generates medical image–mask pairs, and CHeart and Cardiac Mesh Flow already model dynamic anatomy [24, 25, 205, 206]. Merely joining outputs or adding time is therefore not enough.

Independent truth is the scarce resource behind several apparent gaps. An image and generated mask can share the same wrong boundary; a smooth periodic sequence can be static; a large virtual cohort can omit the target disease tail. MedSegFactory’s inspected design is two-dimensional, and processed segmentation or registration references do not become independent motion truth because they are dense [25, 206, 245]. The useful next question is which wrong output the current evaluator would accept. Medical applications sharpen that question by linking the error to a measurement, diagnosis, treatment quantity or validated trial endpoint.

5 Medical Needs and Validation

A medical task starts from the decision that an error could change. The relevant chain is target population, acquisition, generated output, interpretation, action and endpoint. A plausible image is useful only if the required link is preserved. Figure 37 illustrates three selected medical workflows, not the entire taxonomy and not a generic maturity ladder. It complements the correspondence maps: those maps explain meanings, whereas this figure locates validation obligations after a medical use has been chosen.

Selected medical workflows: where validation enters

Context of use -> independent reference -> consequential endpoint

Medical use		Independent reference		Decision / failure
Acquisition / reconstruction Accelerated MRI or CT Measured operator + noise	—	Acquired data + local truth Held-out patients / scanners Reader or quantitative test	—	Lesion / measurement error Acceptance and reacquisition Total time and fallback
Translation / staining MR-only RT or virtual histology Preserve identity and biology	—	Dose / positioning reference Registered tissue + readers Finding-specific disagreement	—	Dose / diagnostic error Artifact or missing structure QA and standard-path fallback
Virtual imaging / device trials Target anatomy and pathology Population + physics + observer	—	External target support Component and endpoint tests Procedural / real comparators	—	Tail endpoint-risk error Population / simulator bias Separate Monte Carlo error

Figure 37: **Selected medical validation paths.** Scope: three use families, not all generative AI or one temporal subtask. Independent references connect generated outputs to a consequential endpoint. Sources: [12, 20, 28, 408–412].

5.1 Acquisition and Reconstruction: Preserve Decision-Relevant Information

Faster acquisition is valuable only if important errors remain acceptable. Reconstruction studies need raw-data splits, realistic operator shifts, strong deterministic and generative comparators, and prespecified lesion or quantitative endpoints. The prospective knee MRI study used 170 participants and six readers to test interchangeability, but lacked surgical truth and used a single vendor [408]. It is a non-diffusion comparator, demonstrating an evaluation design rather than validating newer priors. Data-consistent samples can still differ in poorly observed structures; hallucination work already makes this problem explicit [18, 89]. A useful uncertainty output should identify cases requiring another measurement or fallback, with the acquisition and review burden included.

Translation becomes consequential when a downstream quantity is used. For MR-only radiotherapy, synthetic CT supports dose calculation and positioning, not merely CT-like appearance. An inspected brain RT workflow completed 20/21 cases and exposed an artifact-related fallback associated with hemostatic material [12]. Its findings attach to that workflow; separate atlas-based synthetic-CT quality assurance is not a universal learned-generator certificate [411]. SynthRAD separates image, geometry and dose criteria [50]. This makes a practical research question precise: does a method preserve dose-relevant quantities under the scanner, pathology and implant conditions that cause failures, and can it recognize when it does not?

5.2 Diagnostic Translation: Preserve the Structure Behind the Label

Virtual staining must preserve diagnostic evidence, not just a staining style. Registration, tissue preparation and reference quality affect paired comparisons. UV-PARS reports diagnostic concordance and a blinded three-pathologist

image-quality comparison, but not improved surgical margins or patient outcomes [409]. Recent context-of-use discussion is relevant to validation design, yet the neighbor inspected only at abstract level cannot carry a detailed empirical claim [412]. A useful study separates finding-specific omissions, introduced structures and unreadable cases, tests tissue and site shifts, and retains pathologist review. Averaging these errors into an attractive perceptual score would erase the reason to perform the translation.

5.3 Virtual Trials: Validate the Population and the Simulator

More virtual patients reduce sampling error only within the assumed model. VICTRE compared mammography systems using procedural breast generation, imaging physics and observers across 2,986 virtual patients, with a human-trial comparison [410]. It was not a learned-generative-AI invention and its demonstrated detection task does not validate malignancy diagnosis or mortality. Limited lesions, motion, views and target populations constrain extrapolation. CAN-FLOW’s conditional anatomy and tail analyses are relevant ingredients, but its inspected healthy static data do not establish diseased dynamic trial cohorts [28]. A learned generator must therefore justify target-population support and endpoint sensitivity in addition to realism. Increasing cohort size cannot eliminate bias in anatomy, physics, observer or sampling assumptions.

Patient-specific simulation needs more than a patient-shaped object. Virtual-heart ablation work includes retrospective and small prospective evidence, but used averaged electrophysiological parameters in parts of the workflow [413]. Dynamic geometry alone does not identify individual material, conduction or treatment-response parameters. Likewise, pose-conditioned ultrasound generation does not determine contact force, tissue deformation or scanner-setting effects [26, 127]. Appropriate intermediate targets are independently checked local motion, measured action–observation relations and validated endpoint changes. The stronger claim of treatment or policy validity should wait for the additional measurements that make it identifiable.

Validation guidance must match the model’s role. FDA credibility guidance concerns computational modeling and simulation within its stated physics/mechanistic scope; it is not blanket approval guidance for all GenAI [414]. DECIDE-AI addresses early live clinical decision-support evaluation, not generic offline synthesis [415]. Both motivate explicit context of use, risk and workflow reporting, but checklist compliance does not establish model quality. The study should name which decision is being supported, what error is tolerable, which independent reference can detect it and what happens when the system declines or fails.

6 Research Opportunities and Limitations

6.1 Choose the Missing Evidence, Not the Fashionable Label

A **promising topic joins a real bottleneck to a newly testable capability**. A crowded keyword can conceal an important unresolved endpoint; a rarely named combination may be unidentifiable, unnecessary or already known under different words. Figure 38 separates literature novelty from the evidence needed to test a useful improvement. Before calling something a gap, distinguish a missing implementation, missing independent validation, unsupported population, absent demand and unavailable reference. These lead to different projects. The four cases below are explicitly medical-oriented hypotheses, not an exhaustive agenda for GenAI, a ranking of all 27 subitems or a prediction of publication acceptance.

From a literature gap to a decisive research pilot

General research-selection workflow | Four medical cases are developed in the text

1 Specify the need

Population and input
Required invariant
Decision-changing error
Available independent truth

Can this error be measured?

No -> narrower claim
or build a validation resource

2 Verify the increment

Nearest work + synonyms
What already exists?
Missing capability or evidence?
Strong conventional baseline

Is the increment testable?

No -> revise the question
not just the model name

3 Run a bounded test

Freeze split and endpoint
Include plausible wrong controls
Match data and compute access
Prespecify a stopping rule

Does the pilot change the consequential result?

No -> stop or narrow
Yes -> ablate, then scale

Figure 38: **General evidence-to-project decision workflow**. Scope: research selection across tasks; this is not a capability or maturity map. The four medical cases are separate hypotheses in the text and Table 4. A failed gate implies a narrower claim, a validation contribution or stopping, not a post-hoc weaker endpoint.

Choose according to the validation resource that is actually available. Raw measurements and reliable decisions favor a selective-reconstruction study; independent local motion favors coupled dynamics; target-disease support and an endpoint-validated simulator favor cohort research; safely tracked actions favor an observation/control study. Without these resources, a narrower benchmark or evaluator can be more useful than another large generator. General ML contributions still require transferable capability tests; a use-specific engineering gain and a clinical workflow gain are different contributions, and neither should inherit the other’s evidence.

6.2 Four Bounded Medical-Oriented Hypotheses

O1: selective reconstruction or translation with decision-calibrated fallback. DPS, hallucination metrics, context-specific evaluation and active measurement selection already exist [18, 19, 109, 412]. Steerable Conditional Diffusion already adapts priors from measurements; DynamicDPS evaluates low-field MRI tissue-volume error; HalluGen supplies controlled hallucinations and detector benchmarks [416–418]. A generic OOD adapter or hallucination detector is therefore not the gap. The narrower hypothesis is that acceptance, rejection or extra acquisition can lower a prespecified decision loss at acceptable total burden. For acceptance a and decision loss L , report

$$\text{coverage} = \mathbb{E}[a], \quad \text{risk} = \frac{\mathbb{E}[aL]}{\mathbb{E}[a]}, \quad \text{burden} = \mathbb{E}[\text{time} + \lambda \text{ fallback cost}], \quad (3)$$

with nonzero acceptance, held-out calibration, explicit units and a justified trade-off parameter. Freeze patient splits and decision margins before testing difficult pathology and acquisition shifts. Stop if only similarity improves, dan-

gerous errors remain unrecognized, or fallback erases the benefit. The claimed increment is decision reliability under a defined shift, not uncertainty visualization alone.

O2: joint anatomy–observation generation grounded in independent local truth. JointNet, MedSegFactory and AortaDiff already couple outputs; CHeart and Cardiac Mesh Flow already model dynamic anatomy [24, 25, 49, 205, 206]. MIVAE also learns cardiac image/mesh representations with cross-domain reconstruction and external feature tests [245]. Thus an extra branch, mesh or temporal dimension does not suffice. Test whether coupling improves a local quantity measured outside the generating annotation pipeline. Compare serial generation, a joint image–mask model and a non-generative measurement baseline; use geometry only when the endpoint needs it. Static outputs, mismatched conditions and globally volume-matched but locally wrong anatomy should defeat a weak evaluator. If the available reference cannot distinguish these controls, build validation data first. If image–mask outputs suffice, forcing a mesh adds representation cost without demonstrated benefit.

O3: target-population tail-risk credibility for virtual cohorts. Conditional anatomy and tails, graph-based cohorts and virtual imaging endpoints already have direct neighbors [28, 410, 419, 420]. Hierarchical credibility work already separates cohort, patient, device and clinician submodels [421]; population-aware validation is not an invention of this agenda. The narrower target is preserving an external endpoint-risk distribution under population and parameter uncertainty. Define disease strata and support first, then compare resampling, PCA, procedural and learned cohorts through the same appropriately validated simulator. Evaluate $R_T = \mathbb{E}_{z \sim P_T}[L(z)]$, including tail error, population uncertainty and Monte Carlo error separately. Weighting cannot create absent disease support, and more samples cannot remove simulator bias. Stop the credible-trial claim if the target population or endpoint validity is unknown; a bounded anatomy-generation result can remain useful.

O4: action-valid observation models beyond recorded trajectories. EchoWorld and SonoWorld are already direct ultrasound neighbors [26, 127]; this is not a proposal for the first medical world model. Test instead whether generated observations preserve action rankings under independently tracked held-out movements and improve control outside replay. Start on a safe phantom, comparing replay, nearest-neighbor retrieval, interpolation and a suitable geometric baseline. Vary contact force, deformation and scanner settings, and separate observation fidelity, action choice and closed-loop success. A real-acquisition transfer claim needs an appropriate approved platform and independent measurements. Stop if gains reduce to smoothness, replay or a policy exploiting simulator artifacts. Predictive autoregression is not treatment-effect identification.

Table 4 turns the four cases into project-entry tests. It is deliberately a resource-and-endpoint guide rather than a second task taxonomy; each case combines several of the preceding coordinates.

Table 4: **Four medical-oriented opportunity tests.** Scope: the hypotheses in Section 6, not all of generative AI. Each requires fresh nearest-neighbor verification before a novelty claim.

Case	Required resource	Decisive comparison	Stop or narrow when
O1	Raw measurements or aligned truth; a defined decision.	Risk–coverage–burden versus strong deterministic and generative methods.	Rejection is uncalibrated or fallback removes the gain.
O2	Independent local geometry/motion reference.	Joint versus serial generation; static and locally wrong controls.	Shared labels cannot detect shared errors.
O3	Supported target disease population; validated endpoint simulator.	Tail risk versus resampling, PCA, procedural and learned cohorts.	Population support or endpoint validity is absent.
O4	Safe platform; independently tracked, held-out actions.	Real action success versus replay, retrieval and interpolation.	Benefits are replay-only or artifact exploitation.

6.3 Limitations and Next Decisions

The survey cannot certify an empty field, a best topic or clinical readiness. The query misses differently named work and includes background mentions; the historical audit and sensitivity tests do not validate all labels or the 2026 extension. Purposive selection favors accessible and salient work. Single-AI extraction lacks independent expert adjudication, and abstract/indexed access is weaker than locally inspected passages. Timeline landmarks are not a citation ranking, while the four opportunities are untested hypotheses. A candidate project’s nearest-neighbor search must be refreshed at design freeze, especially for recent preprints. No field-wide absence, global first, regulatory assessment or acceptance-probability claim follows here.

A small decisive pilot is the next useful output. Write the target population, available input, required invariant, independent reference, comparator and stopping condition on one page before model training. First verify that the reference rejects plausible wrong controls; then test the smallest system capable of the proposed increment. Broaden only after this gate passes. A negative pilot can justify a narrower claim, a better evaluator or a different resource allocation. Relaxing the endpoint after seeing the result would instead remove the reason the study mattered.

7 Conclusion

The eight coordinates and 27 histories separate what generation does from its conditions, resources, representations and mechanism. This makes both reuse and disagreement easier to understand: changing a model family need not change the task, while the same operation can require very different truth in CV and medicine. The descriptive corpus shows shifts toward action-related language and newer mechanisms; it does not select a research topic by itself. The stronger basis is a concrete unmet endpoint with independent truth and a comparator capable of meeting the same need. For medical applications, selective decision reliability, independently grounded dynamics, target-population support and action-valid observations are useful bounded tests. The next contribution should be the smallest verified capability that changes a consequential result, not the largest collection of generated outputs.

Transparency and Reproducibility

This is a Codex-assisted research draft, not an independently adjudicated clinical review. Retrieval, passage extraction, coding, synthesis and writing used the current Codex system; expert verification is required before submission or clinical interpretation. No new patient experiment, training, ethics approval or prospective registration is claimed. The archive includes source identities, inspection locations, coding rules, analysis tables, editable figure sources and compilation instructions. Third-party full-text PDFs are not redistributed. Journal rankings and popularity are not evidence grades. Limitations in Section 6.3 apply throughout.

References

- [1] Florinel-Alin Croitoru, Vlad Hondru, Radu Tudor Ionescu, and Mubarak Shah. Diffusion Models in Vision: A Survey. arXiv:2209.04747, 2022.
- [2] Xuanru Zhou, Cheng Li, Shuqiang Wang, Ye Li, Tao Tan, Hairong Zheng, and Shanshan Wang. Generative Artificial Intelligence in Medical Imaging: Foundations, Progress, and Clinical Translation. arXiv:2508.09177, 2025.
- [3] Fengming Lin. Vision Language Models: A Survey of 26K Papers. arXiv:2510.09586, 2025.
- [4] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. arXiv:1312.6114, 2013.
- [5] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G. Dimakis. Compressed Sensing using Generative Models. arXiv:1703.03208, 2017.
- [6] Utkarsh Ojha, Yijun Li, Jingwan Lu, Alexei A. Efros, Yong Jae Lee, Eli Shechtman, and Richard Zhang. Few-shot Image Generation via Cross-domain Correspondence. arXiv:2104.06820, 2021.
- [7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752, 2021.
- [8] Dave Zhenyu Chen, Yawar Siddiqui, Hsin-Ying Lee, Sergey Tulyakov, and Matthias Nießner. Text2Tex: Text-driven Texture Synthesis via Diffusion Models. arXiv:2303.11396, 2023.
- [9] Pengfei Guo, Can Zhao, Dong Yang, Ziyue Xu, Vishwesh Nath, Yucheng Tang, Benjamin Simon, Mason Belue, Stephanie Harmon, Baris Turkbey, and Daguang Xu. MAISI: Medical AI for Synthetic Imaging. arXiv:2409.11169, 2024.
- [10] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. arXiv:1611.07004, 2016.
- [11] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXiv:1703.10593, 2017.
- [12] Minna Lerner, Joakim Medin, Christian Jamtheim Gustafsson, Sara Alkner, and Lars E Olsson. Prospective Clinical Feasibility Study for MRI-Only Brain Radiotherapy. Primary source, 2022.
- [13] Yair Rivenson, Hongda Wang, Zhensong Wei, Yibo Zhang, Harun Gunaydin, and Aydogan Ozcan. Deep learning-based virtual histology staining using auto-fluorescence of label-free tissue. arXiv:1803.11293, 2018.

- [14] Tim Brooks, Aleksander Holynski, and Alexei A. Efros. InstructPix2Pix: Learning to Follow Image Editing Instructions. arXiv:2211.09800, 2022.
- [15] Karim Kadry, Shreya Gupta, Farhad R. Nezami, and Elazer R. Edelman. Probing the limits and capabilities of diffusion models for the anatomic editing of digital twins. arXiv:2401.00247, 2024.
- [16] Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Palette: Image-to-Image Diffusion Models. arXiv:2111.05826, 2021.
- [17] Bahjat Kavar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising Diffusion Restoration Models. arXiv:2201.11793, 2022.
- [18] Matthew Tivnan, Siyeop Yoon, Zhenhong Chen, Xiang Li, Dufan Wu, and Quanzheng Li. Hallucination Index: An Image Quality Metric for Generative Reconstruction Models. arXiv:2407.12780, 2024.
- [19] Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion Posterior Sampling for General Noisy Inverse Problems. arXiv:2209.14687, 2022.
- [20] Yang Song, Liye Shen, Lei Xing, and Stefano Ermon. Solving Inverse Problems in Medical Imaging with Score-Based Generative Models. arXiv:2111.08005, 2021.
- [21] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J. Fleet. Video Diffusion Models. arXiv:2204.03458, 2022.
- [22] Jake Bruce, Michael Dennis, Ashley Edwards, Jack Parker-Holder, Yuge Shi, Edward Hughes, Matthew Lai, Aditi Mavalankar, Richie Steigerwald, Chris Apps, Yusuf Aytar, Sarah Bechtle, Feryal Behbahani, Stephanie Chan, Nicolas Heess, Lucy Gonzalez, Simon Osindero, Sherjil Ozair, Scott Reed, Jingwei Zhang, Konrad Zolna, Jeff Clune, Nando de Freitas, Satinder Singh, and Tim Rocktäschel. Genie: Generative Interactive Environments. arXiv:2402.15391, 2024.
- [23] Sherwin Bahmani, Ivan Skorokhodov, Victor Rong, Gordon Wetzstein, Leonidas Guibas, Peter Wonka, Sergey Tulyakov, Jeong Joon Park, Andrea Tagliasacchi, and David B. Lindell. 4D-fy: Text-to-4D Generation Using Hybrid Score Distillation Sampling. arXiv:2311.17984, 2023.
- [24] Mengyun Qiao, Shuo Wang, Huaqi Qiu, Antonio de Marvao, Declan P. O'Regan, Daniel Rueckert, and Wenjia Bai. CHeart: A Conditional Spatio-Temporal Generative Model for Cardiac Anatomy. arXiv:2301.13098, 2023.
- [25] Qiang Ma, Qingjie Meng, Mengyun Qiao, Paul M. Matthews, Declan P. O'Regan, and Wenjia Bai. Cardiac Mesh Flow: One-Step Generation of 3D+t Cardiac Four-Chamber Meshes via Flow Matching. arXiv:2605.01884, 2026.
- [26] Zefan Yang, Zheyuan Zhang, Boris Maihe, Yikang Liu, Hang Yu, Ankush Mukherjee, Xiao Chen, Pingkun Yan, Terrence Chen, and Shanhui Sun. SonoWorld: Towards Interactive Pose-Controllable Ultrasound World Models. arXiv:2610.19327, 2026.
- [27] Can Zhao, Pengfei Guo, Dong Yang, Yucheng Tang, Yufan He, Benjamin Simon, Mason Belue, Stephanie Harmon, Baris Turkbey, and Daguang Xu. MAISI-v2: Accelerated 3D High-Resolution Medical Image Synthesis with Rectified Flow and Region-specific Contrastive Loss. arXiv:2508.05772, 2025.
- [28] Konstantinos Kevopoulos, Beatrice Moscoloni, Benjamin Alheit, Cameron Beeche, Julio A. Chirinos, Alexander Heinlein, and Mathias Peirlinck. Flow-based conditional cardiac anatomy generation for virtual cohorts. arXiv:2608.09460, 2026.
- [29] Fanwei Kong, Sascha Stocker, Perry S. Choi, Michael Ma, Daniel B. Ennis, and Alison Marsden. SDF4CHD: Generative Modeling of Cardiac Anatomies with Congenital Heart Defects. arXiv:2311.00332, 2023.
- [30] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large Scale GAN Training for High Fidelity Natural Image Synthesis. arXiv:1809.11096, 2018.
- [31] Tero Karras, Samuli Laine, and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. arXiv:1812.04948, 2018.
- [32] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and Improving the Image Quality of StyleGAN. arXiv:1912.04958, 2019.
- [33] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. arXiv:2006.11239, 2020.
- [34] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. arXiv:2011.13456, 2020.
- [35] Patrick Esser, Robin Rombach, and Björn Ommer. Taming Transformers for High-Resolution Image Synthesis. arXiv:2012.09841, 2020.

- [36] Alex Nichol and Prafulla Dhariwal. Improved Denoising Diffusion Probabilistic Models. arXiv:2102.09672, 2021.
- [37] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-Shot Text-to-Image Generation. arXiv:2102.12092, 2021.
- [38] Prafulla Dhariwal and Alex Nichol. Diffusion Models Beat GANs on Image Synthesis. arXiv:2105.05233, 2021.
- [39] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. arXiv:2112.10741, 2021.
- [40] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical Text-Conditional Image Generation with CLIP Latents. arXiv:2204.06125, 2022.
- [41] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J Fleet, and Mohammad Norouzi. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. arXiv:2205.11487, 2022.
- [42] Jiahui Yu, Yuanzhong Xu, Jing Yu Koh, Thang Luong, Gunjan Baid, Zirui Wang, Vijay Vasudevan, Alexander Ku, Yinfei Yang, Burcu Karagol Ayan, Ben Hutchinson, Wei Han, Zarana Parekh, Xin Li, Han Zhang, Jason Baldridge, and Yonghui Wu. Scaling Autoregressive Models for Content-Rich Text-to-Image Generation. arXiv:2206.10789, 2022.
- [43] Firas Khader, Gustav Mueller-Franzes, Soroosh Tayebi Arasteh, Tianyu Han, Christoph Haarbuerger, Maximilian Schulze-Hagen, Philipp Schad, Sandy Engelhardt, Bettina Baessler, Sebastian Foersch, Johannes Stegmaier, Christiane Kuhl, Sven Nebelung, Jakob Nikolas Kather, and Daniel Truhn. Medical Diffusion: Denoising Diffusion Probabilistic Models for 3D Medical Image Generation. arXiv:2211.03364, 2022.
- [44] William Peebles and Saining Xie. Scalable Diffusion Models with Transformers. arXiv:2212.09748, 2022.
- [45] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency Models. arXiv:2303.01469, 2023.
- [46] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. arXiv:2307.01952, 2023.
- [47] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, Kyle Lacey, Alex Goodwin, Yannik Marek, and Robin Rombach. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. arXiv:2403.03206, 2024.
- [48] Bing Zhao, Chenfei Wu, Deqing Li, Hao Meng, Jiahao Li, Jie Zhang, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kuan Cao, Kun Yan, Liang Peng, Lihan Jiang, Niantong Li, Ningyuan Tang, Shengming Yin, Tianhe Wu, Xiao Xu, Xiaoyue Chen, Xihua Wang, Yan Shu, Yanran Zhang, Yi Wang, Yilei Chen, Ying Ba, Yixian Xu, Yujia Wu, Yuxiang Chen, Zecheng Tang, Zekai Zhang, Zhendong Wang, Zihao Liu, Zikai Zhou, An Yang, Chen Cheng, Chenxu Lv, Dayiheng Liu, Fan Zhou, Hantian Xiong, Hongzhu Shi, Hu Wei, Huihong Zhao, Ivy Liu, Jianwei Zhang, Jiawei Zhang, Kai Chen, Kang He, Levon Xue, Lin Qu, Linhan Tang, Luwen Feng, Minggang Wu, Minmin Sun, Na Ni, Rui Men, Shuai Bai, Sishou Zheng, Tao Lan, Tianqi Zhang, Tingkun Wen, Wei Wang, Weixu Qiao, Weiye Lu, Wenmeng Zhou, Xiaodong Deng, Xiaoxiao Xu, Xinlei Fang, Xionghui Chen, Yanan Wang, Yang Fan, Yichang Zhang, Yixuan Xu, Yu Wu, Zhiyuan Ma, and Zhizhi Cai. Qwen-Image-2.0 Technical Report. arXiv:2605.10730, 2026.
- [49] Yuxuan Ou, Ning Bi, Jiazhen Pan, Jiancheng Yang, Boliang Yu, Usama Zidan, Regent Lee, and Vicente Grau. AortaDiff: A Unified Multitask Diffusion Framework For Contrast-Free AAA Imaging. arXiv:2510.01498, 2025.
- [50] Viktor Rogowski, Maarten L. Terpstra, Niklas Wahl, Florian Kamp, Erik van der Bijl, Arthur Jr. Galapon, Christopher Kurz, Bowen Xin, Zhengxiang Sun, Hollie Min, Gregg Belous, Jason Dowling, Yan Xia, Siyuan Mei, Fuxin Fan, Arthur Longuefosse, Javier Sequeiro Gonzalez, Miguel Diaz Benito, Alvaro Garcia Martin, Fabien Baldacci, Valentin Boussot, Cédric Hémon, Jean-Claude Nunes, Jean-Louis Dillenseger, Zhiyuan Zhang, Jinghua Cai, Han Bing, Tan Zuopeng, Ricardo Brioso, Daniele Loiacono, Guillaume Landry, Adrian Thummerer, and Matteo Maspero. Generating synthetic computed tomography for radiotherapy: SynthRAD2025 challenge report. arXiv:2605.13555, 2026.
- [51] Salman Ul Hassan Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, Erkut Erdem, and Tolga Çukur. Image Synthesis in Multi-Contrast MRI with Conditional Generative Adversarial Networks. arXiv:1802.01221, 2018.

- [52] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal Unsupervised Image-to-Image Translation. arXiv:1804.04732, 2018.
- [53] Karim Armanious, Chenming Jiang, Marc Fischer, Thomas Küstner, Konstantin Nikolaou, Sergios Gatidis, and Bin Yang. MedGAN: Medical Image Translation using GANs. arXiv:1806.06397, 2018.
- [54] Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Kumar Singh, and Ming-Hsuan Yang. Diverse Image-to-Image Translation via Disentangled Representations. arXiv:1808.00948, 2018.
- [55] Heran Yang, Jian Sun, Aaron Carass, Can Zhao, Junghoon Lee, Zongben Xu, and Jerry Prince. Unpaired Brain MR-to-CT Synthesis using a Structure-Constrained CycleGAN. arXiv:1809.04536, 2018.
- [56] Ming-Yu Liu, Xun Huang, Arun Mallya, Tero Karras, Timo Aila, Jaakko Lehtinen, and Jan Kautz. Few-Shot Unsupervised Image-to-Image Translation. arXiv:1905.01723, 2019.
- [57] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation. arXiv:1907.10830, 2019.
- [58] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. StarGAN v2: Diverse Image Synthesis for Multiple Domains. arXiv:1912.01865, 2019.
- [59] Runfa Chen, Wenbing Huang, Binghui Huang, Fuchun Sun, and Bin Fang. Reusing Discriminators for Encoding: Towards Unsupervised Image-to-Image Translation. arXiv:2003.00273, 2020.
- [60] Kyungjune Baek, Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Hyunjeong Shim. Rethinking the Truly Unsupervised Image-to-Image Translation. arXiv:2006.06500, 2020.
- [61] Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive Learning for Unpaired Image-to-Image Translation. arXiv:2007.15651, 2020.
- [62] Hiroshi Sasaki, Chris G. Willcocks, and Toby P. Breckon. UNIT-DDPM: UNpaired Image Translation with Denoising Diffusion Probabilistic Models. arXiv:2104.05358, 2021.
- [63] Junlin Han, Mehrdad Sholeby, Lars Petersson, and Mohammad Ali Armin. Dual Contrastive Learning for Unsupervised Image-to-Image Translation. arXiv:2104.07689, 2021.
- [64] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models. arXiv:2108.02938, 2021.
- [65] Xuan Su, Jiaming Song, Chenlin Meng, and Stefano Ermon. Dual Diffusion Implicit Bridges for Image-to-Image Translation. arXiv:2203.08382, 2022.
- [66] Muzaffer Özbey, Onat Dalmaz, Salman UH Dar, Hasan A Bedel, Şaban Öztürk, Alper Güngör, and Tolga Çukur. Unsupervised Medical Image Translation with Adversarial Diffusion Models. arXiv:2207.08208, 2022.
- [67] Gihyun Kwon and Jong Chul Ye. Diffusion-based Image Translation using Disentangled Style and Content Representation. arXiv:2209.15264, 2022.
- [68] Beomsu Kim, Gihyun Kwon, Kwanyoung Kim, and Jong Chul Ye. Unpaired Image-to-Image Translation via Neural Schrödinger Bridge. arXiv:2305.15086, 2023.
- [69] Gaurav Parmar, Taesung Park, Srinivasa Narasimhan, and Jun-Yan Zhu. One-Step Image Translation with Text-to-Image Models. arXiv:2403.12036, 2024.
- [70] Yifan Chen, Fei Yin, Hao Chen, Jia Wu, and Chao Li. Contrast-X: A Multi-Modal Contrast Image Synthesis Benchmark and Universal Modality Flow Matching. arXiv:2601.15884, 2026.
- [71] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the Latent Space of GANs for Semantic Face Editing. arXiv:1907.10786, 2019.
- [72] Erik Härkönen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. GANSpace: Discovering Interpretable GAN Controls. arXiv:2004.02546, 2020.
- [73] Rameen Abdal, Peihao Zhu, Niloy Mitra, and Peter Wonka. StyleFlow: Attribute-conditioned Exploration of StyleGAN-Generated Images using Conditional Continuous Normalizing Flows. arXiv:2008.02401, 2020.
- [74] Or Patashnik, Zongze Wu, Eli Shechtman, Daniel Cohen-Or, and Dani Lischinski. StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery. arXiv:2103.17249, 2021.
- [75] Rinon Gal, Or Patashnik, Haggai Maron, Gal Chechik, and Daniel Cohen-Or. StyleGAN-NADA: CLIP-Guided Domain Adaptation of Image Generators. arXiv:2108.00946, 2021.
- [76] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations. arXiv:2108.01073, 2021.

- [77] Omri Avrahami, Dani Lischinski, and Ohad Fried. Blended Diffusion for Text-driven Editing of Natural Images. arXiv:2111.14818, 2021.
- [78] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-Prompt Image Editing with Cross Attention Control. arXiv:2208.01626, 2022.
- [79] Bahjat Kavar, Shiran Zada, Oran Lang, Omer Tov, Huiwen Chang, Tali Dekel, Inbar Mosseri, and Michal Irani. Imagic: Text-Based Real Image Editing with Diffusion Models. arXiv:2210.09276, 2022.
- [80] Guillaume Couairon, Jakob Verbeek, Holger Schwenk, and Matthieu Cord. DiffEdit: Diffusion-based semantic image editing with mask guidance. arXiv:2210.11427, 2022.
- [81] Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text Inversion for Editing Real Images using Guided Diffusion Models. arXiv:2211.09794, 2022.
- [82] Narek Tumanyan, Michal Geyer, Shai Bagon, and Tali Dekel. Plug-and-Play Diffusion Features for Text-Driven Image-to-Image Translation. arXiv:2211.12572, 2022.
- [83] Mingdeng Cao, Xintao Wang, Zhongang Qi, Ying Shan, Xiaohu Qie, and Yinqiang Zheng. MasaCtrl: Tuning-Free Mutual Self-Attention Control for Consistent Image Synthesis and Editing. arXiv:2304.08465, 2023.
- [84] Xingang Pan, Ayush Tewari, Thomas Leimkühler, Lingjie Liu, Abhimitra Meka, and Christian Theobalt. Drag Your GAN: Interactive Point-based Manipulation on the Generative Image Manifold. arXiv:2305.10973, 2023.
- [85] Yujun Shi, Chuhui Xue, Jun Hao Liew, Jiachun Pan, Hanshu Yan, Wenqing Zhang, Vincent Y. F. Tan, and Song Bai. DragDiffusion: Harnessing Diffusion Models for Interactive Point-based Image Editing. arXiv:2306.14435, 2023.
- [86] Manuel Brack, Felix Friedrich, Katharina Kornmeier, Linoy Tsaban, Patrick Schramowski, Kristian Kersting, and Apolinário Passos. LEDITS++: Limitless Image Editing using Text-to-Image Models. arXiv:2311.16711, 2023.
- [87] Malek Ben Alaya, Daniel M. Lang, Benedikt Wiestler, Julia A. Schnabel, and Cosmin I. Bercea. MedEdit: Counterfactual Diffusion-based Image Editing on Brain MRI. arXiv:2407.15270, 2024.
- [88] Black Forest Labs, Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, Sumith Kulal, Kyle Lacey, Yam Levi, Cheng Li, Dominik Lorenz, Jonas Müller, Dustin Podell, Robin Rombach, Harry Saini, Axel Sauer, and Luke Smith. FLUX.1 Kontext: Flow Matching for In-Context Image Generation and Editing in Latent Space. arXiv:2506.15742, 2025.
- [89] Sayantan Bhadra, Varun A. Kelkar, Frank J. Brooks, and Mark A. Anastasio. On hallucinations in tomographic image reconstruction. arXiv:2012.00646, 2020.
- [90] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S. Huang. Generative Image Inpainting with Contextual Attention. arXiv:1801.07892, 2018.
- [91] Guilin Liu, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image Inpainting for Irregular Holes Using Partial Convolutions. arXiv:1804.07723, 2018.
- [92] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. arXiv:1809.00219, 2018.
- [93] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Z. Qureshi, and Mehran Ebrahimi. EdgeConnect: Generative Image Inpainting with Adversarial Edge Learning. arXiv:1901.00212, 2019.
- [94] Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. Pluralistic Image Completion. arXiv:1903.04227, 2019.
- [95] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. SRFlow: Learning the Super-Resolution Space with Normalizing Flow. arXiv:2006.14200, 2020.
- [96] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards Real-World Blind Face Restoration with Generative Facial Prior. arXiv:2101.04061, 2021.
- [97] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Image Super-Resolution via Iterative Refinement. arXiv:2104.07636, 2021.
- [98] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. arXiv:2107.10833, 2021.
- [99] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust Large Mask Inpainting with Fourier Convolutions. arXiv:2109.07161, 2021.

- [100] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting using Denoising Diffusion Probabilistic Models. arXiv:2201.09865, 2022.
- [101] Shangchen Zhou, Kelvin C. K. Chan, Chongyi Li, and Chen Change Loy. Towards Robust Blind Face Restoration with Codebook Lookup Transformer. arXiv:2206.11253, 2022.
- [102] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C. K. Chan, and Chen Change Loy. Exploiting Diffusion Prior for Real-World Image Super-Resolution. arXiv:2305.07015, 2023.
- [103] Xinqi Lin, Jingwen He, Ziyang Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. DiffBIR: Towards Blind Image Restoration with Generative Diffusion Prior. arXiv:2308.15070, 2023.
- [104] Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C. Kot, and Bihan Wen. SinSR: Diffusion-Based Image Super-Resolution in a Single Step. arXiv:2311.14760, 2023.
- [105] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution. arXiv:2311.16518, 2023.
- [106] Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration In the Wild. arXiv:2401.13627, 2024.
- [107] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-Step Effective Diffusion Network for Real-World Image Super-Resolution. arXiv:2406.08177, 2024.
- [108] Yoonseok Choi, Eun-Gyu Ha, Daniel Kim, Mohammed A. Al-masni, Ming-Hsuan Yang, and Dong-Hyun Kim. DRIFT: Difficulty-aware Rectified Flows for Through-plane MRI Super-Resolution. arXiv:2607.16649, 2026.
- [109] Luis Barba, Johannes Kirschner, Tomas Aidukas, Manuel Guizar-Sicairos, and Benjamín Béjar. Diffusion Active Learning: Towards Data-Driven Experimental Design in Computed Tomography. arXiv:2504.03491, 2025.
- [110] Maya Kabkab, Pouya Samangouei, and Rama Chellappa. Task-Aware Compressed Sensing with Generative Adversarial Networks. arXiv:1802.01284, 2018.
- [111] Manik Dhar, Aditya Grover, and Stefano Ermon. Modeling Sparse Deviations for Compressed Sensing using Generative Models. arXiv:1807.01442, 2018.
- [112] Zhongnian Li, Tao Zhang, Peng Wan, and Daoqiang Zhang. SEGAN: Structure-Enhanced Generative Adversarial Network for Compressed Sensing MRI Reconstruction. arXiv:1902.06455, 2019.
- [113] Kyung-Su Kim, Jung Hyun Lee, and Eunho Yang. Compressed Sensing via Measurement-Conditional Generative Models. arXiv:2007.00873, 2020.
- [114] Ajil Jalal, Marius Arvinte, Giannis Daras, Eric Price, Alexandros G. Dimakis, and Jonathan I. Tamir. Robust Compressed Sensing MRI with Deep Generative Priors. arXiv:2108.01368, 2021.
- [115] Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated MRI. arXiv:2110.05243, 2021.
- [116] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving Diffusion Models for Inverse Problems using Manifold Constraints. arXiv:2206.00941, 2022.
- [117] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model. arXiv:2212.00490, 2022.
- [118] Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed Diffusion Sampler for Accelerating Large-Scale Inverse Problems. arXiv:2303.05754, 2023.
- [119] Morteza Mardani, Jiaming Song, Jan Kautz, and Arash Vahdat. A Variational Perspective on Solving Inverse Problems with Diffusion Models. arXiv:2305.04391, 2023.
- [120] Benjamin Boys, Mark Girolami, Jakiw Pidstrigach, Sebastian Reich, Alan Mosca, and O. Deniz Akyildiz. Tweedie Moment Projected Diffusions For Inverse Problems. arXiv:2310.06721, 2023.
- [121] Shudong Li, Xiao Jiang, Matthew Tivnan, Grace J. Gang, Yuan Shen, and J. Webster Stayman. CT Reconstruction using Diffusion Posterior Sampling conditioned on a Nonlinear Measurement Model. arXiv:2312.01464, 2023.
- [122] Asad Aali, Giannis Daras, Brett Levac, Sidharth Kumar, Alexandros G. Dimakis, and Jonathan I. Tamir. Ambient Diffusion Posterior Sampling: Solving Inverse Problems with Diffusion Models Trained on Corrupted Data. arXiv:2403.08728, 2024.

- [123] Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L. Bouman. Principled Probabilistic Imaging using Diffusion Models as Plug-and-Play Priors. *arXiv:2405.18782*, 2024.
- [124] Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving Diffusion Inverse Problem Solving with Decoupled Noise Annealing. *arXiv:2407.01521*, 2024.
- [125] Yiming Yang, Xiaoyuan Cheng, Yi He, Kaiyu Li, Wenxuan Yuan, and Zhuo Sun. Outlier-robust Diffusion Posterior Sampling for Bayesian Inverse Problems. *arXiv:2602.02045*, 2026.
- [126] Florian F rnrohr and Reinhard Heckel. Piecewise Dynamic Diffusion Regularization for Reconstruction of Cardiac Cine MRI. *arXiv:2607.03299*, 2026.
- [127] Yang Yue, Yulin Wang, Haojun Jiang, Pan Liu, Shiji Song, and Gao Huang. EchoWorld: Learning Motion-Aware World Models for Echocardiography Probe Guidance. Primary source, 2025.
- [128] Alex X. Lee, Richard Zhang, Frederik Ebert, Pieter Abbeel, Chelsea Finn, and Sergey Levine. Stochastic Adversarial Video Prediction. *arXiv:1804.01523*, 2018.
- [129] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Guilin Liu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. Video-to-Video Synthesis. *arXiv:1808.06601*, 2018.
- [130] Aidan Clark, Jeff Donahue, and Karen Simonyan. Adversarial Video Generation on Complex Datasets. *arXiv:1907.06571*, 2019.
- [131] Wilson Yan, Yunzhi Zhang, Pieter Abbeel, and Aravind Srinivas. VideoGPT: Video Generation using VQ-VAE and Transformers. *arXiv:2104.10157*, 2021.
- [132] Mingyuan Zhang, Zhongang Cai, Liang Pan, Fangzhou Hong, Xinying Guo, Lei Yang, and Ziwei Liu. Motion-Diffuse: Text-Driven Human Motion Generation with Diffusion Model. *arXiv:2208.15001*, 2022.
- [133] Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, Devi Parikh, Sonal Gupta, and Yaniv Taigman. Make-A-Video: Text-to-Video Generation without Text-Video Data. *arXiv:2209.14792*, 2022.
- [134] Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit H. Bermano. Human Motion Diffusion Model. *arXiv:2209.14916*, 2022.
- [135] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David J. Fleet, and Tim Salimans. Imagen Video: High Definition Video Generation with Diffusion Models. *arXiv:2210.02303*, 2022.
- [136] Ruben Villegas, Mohammad Babaeizadeh, Pieter-Jan Kindermans, Hernan Moraldo, Han Zhang, Mohammad Taghi Saffar, Santiago Castro, Julius Kunze, and Dumitru Erhan. Phenaki: Variable Length Video Generation From Open Domain Textual Description. *arXiv:2210.02399*, 2022.
- [137] Jay Zhangjie Wu, Yixiao Ge, Xintao Wang, Weixian Lei, Yuchao Gu, Yufei Shi, Wynne Hsu, Ying Shan, Xiaohu Qie, and Mike Zheng Shou. Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation. *arXiv:2212.11565*, 2022.
- [138] Jianrong Zhang, Yangsong Zhang, Xiaodong Cun, Shaoli Huang, Yong Zhang, Hongwei Zhao, Hongtao Lu, and Xi Shen. T2M-GPT: Generating Human Motion from Textual Descriptions with Discrete Representations. *arXiv:2301.06052*, 2023.
- [139] Levon Khachatryan, Andranik Movsisyan, Vahram Tadevosyan, Roberto Henschel, Zhangyang Wang, Shant Navasardyan, and Humphrey Shi. Text2Video-Zero: Text-to-Image Diffusion Models are Zero-Shot Video Generators. *arXiv:2303.13439*, 2023.
- [140] Andreas Blattmann, Robin Rombach, Huan Ling, Tim Dockhorn, Seung Wook Kim, Sanja Fidler, and Karsten Kreis. Align your Latents: High-Resolution Video Synthesis with Latent Diffusion Models. *arXiv:2304.08818*, 2023.
- [141] Yuwei Guo, Ceyuan Yang, Anyi Rao, Zhengyang Liang, Yaohui Wang, Yu Qiao, Maneesh Agrawala, Dahua Lin, and Bo Dai. AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning. *arXiv:2307.04725*, 2023.
- [142] Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, Varun Jampani, and Robin Rombach. Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets. *arXiv:2311.15127*, 2023.
- [143] Omer Bar-Tal, Hila Chefer, Omer Tov, Charles Herrmann, Roni Paiss, Shiran Zada, Ariel Ephrat, Junhwa Hur, Guanghui Liu, Amit Raj, Yuanzhen Li, Michael Rubinstein, Tomer Michaeli, Oliver Wang, Deqing Sun, Tali Dekel, and Inbar Mosseri. Lumiere: A Space-Time Diffusion Model for Video Generation. *arXiv:2401.12945*, 2024.

- [144] Zhuoyi Yang, Jiayan Teng, Wendi Zheng, Ming Ding, Shiyu Huang, Jiazheng Xu, Yuanming Yang, Wenyi Hong, Xiaohan Zhang, Guanyu Feng, Da Yin, Yuxuan Zhang, Weihang Wang, Yean Cheng, Bin Xu, Xiaotao Gu, Yuxiao Dong, and Jie Tang. CogVideoX: Text-to-Video Diffusion Models with An Expert Transformer. arXiv:2408.06072, 2024.
- [145] Marvin Seyfarth, Sarah Kaye Müller, Arman Ghanaat, Isabelle Ayy, Fabian Fastenrath, Philipp Wild, Alexander Hertel, Theano Papavassiliou, Salman Ul Hassan Dar, and Sandy Engelhardt. CardioDiT: Latent Diffusion Transformers for 4D Cardiac MRI Synthesis. arXiv:2603.25194, 2026.
- [146] Team Seedance, De Chen, Liyang Chen, Xin Chen, Ying Chen, Zhuo Chen, Zhuowei Chen, Feng Cheng, Tianheng Cheng, Yufeng Cheng, Mojie Chi, Xuyan Chi, Jian Cong, Qinpeng Cui, Fei Ding, Qide Dong, Yujiao Du, Haojie Duanmu, Junliang Fan, Jiarui Fang, Jing Fang, Zetao Fang, Chengjian Feng, Yu Gao, Diandian Gu, Dong Guo, Hanzhong Guo, Qiushan Guo, Boyang Hao, Hongxiang Hao, Haoxun He, Jiaao He, Qian He, Tuyen Hoang, Heng Hu, Ruqing Hu, Yuxiang Hu, Jiancheng Huang, Weilin Huang, Zhaoyang Huang, Zhongyi Huang, Jishuo Jin, Ming Jing, Ashley Kim, Shanshan Lao, Yichong Leng, Bingchuan Li, Gen Li, Haifeng Li, Huixia Li, Jiashi Li, Ming Li, Xiaojie Li, Xingxing Li, Yameng Li, Yiying Li, Yu Li, Yueyan Li, Chao Liang, Han Liang, Jianzhong Liang, Ying Liang, Wang Liao, J. H. Lien, Shanchuan Lin, Xi Lin, Feng Ling, Yue Ling, Fangfang Liu, Jiawei Liu, Jihao Liu, Jingtuo Liu, Shu Liu, Sichao Liu, Wei Liu, Xue Liu, Zuxi Liu, Ruijie Lu, Lecheng Lyu, Jingting Ma, Tianxiang Ma, Xiaonan Nie, Jingzhe Ning, Junjie Pan, Xitong Pan, Ronggui Peng, Xueqiong Qu, Yuxi Ren, Yuchen Shen, Guang Shi, Lei Shi, Yinglong Song, Fan Sun, Li Sun, Renfei Sun, Wenjing Tang, Boyang Tao, Zirui Tao, Dongliang Wang, Feng Wang, Hulin Wang, Ke Wang, Qingyi Wang, Rui Wang, Shuai Wang, Shulei Wang, Weichen Wang, Xuanda Wang, Yanhui Wang, Yue Wang, Yuping Wang, Yuxuan Wang, Zijie Wang, Ziyu Wang, Guoqiang Wei, Meng Wei, Di Wu, Guohong Wu, Hanjie Wu, Huachao Wu, Jian Wu, Jie Wu, Ruolan Wu, Shaojin Wu, Xiaohu Wu, Xinglong Wu, Yonghui Wu, Ruiqi Xia, Xin Xia, Xuefeng Xiao, Shuang Xu, Bangbang Yang, Jiaqi Yang, Runkai Yang, Tao Yang, Yihang Yang, Zhixian Yang, Ziyang Yang, Fulong Ye, Bingqian Yi, Xing Yin, Yongbin You, Linxiao Yuan, Weihong Zeng, Xuejiao Zeng, Yan Zeng, Siyu Zhai, Zhonghua Zhai, Bowen Zhang, Chenlin Zhang, Heng Zhang, Jun Zhang, Manlin Zhang, Peiyuan Zhang, Shuo Zhang, Xiaohe Zhang, Xiaoying Zhang, Xinyan Zhang, Xinyi Zhang, Yichi Zhang, Zixiang Zhang, Haiyu Zhao, Huating Zhao, Liming Zhao, Yian Zhao, Guangcong Zheng, Jianbin Zheng, Xiaozheng Zheng, Zerong Zheng, Kuan Zhu, and Feilong Zuo. Seedance 2.0: Advancing Video Generation for World Complexity. arXiv:2604.14148, 2026.
- [147] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic Image Synthesis with Spatially-Adaptive Normalization. arXiv:1903.07291, 2019.
- [148] Pierre Chambon, Christian Bluethgen, Jean-Benoit Delbrouck, Rogier Van der Sluijs, Małgorzata Połacin, Juan Manuel Zambrano Chaves, Tanishq Mathew Abraham, Shivanshu Purohit, Curtis P. Langlotz, and Akshay Chaudhari. RoentGen: Vision-Language Foundation Model for Chest X-ray Generation. arXiv:2211.12737, 2022.
- [149] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv:2302.05543, 2023.
- [150] Bo Zhao, Lili Meng, Weidong Yin, and Leonid Sigal. Image Generation from Layout. arXiv:1811.11389, 2018.
- [151] Wenbo Li, Pengchuan Zhang, Lei Zhang, Qiuyuan Huang, Xiaodong He, Siwei Lyu, and Jianfeng Gao. Object-driven Text-to-Image Synthesis via Adversarial Training. arXiv:1902.10740, 2019.
- [152] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. DM-GAN: Dynamic Memory Generative Adversarial Networks for Text-to-Image Synthesis. arXiv:1904.01310, 2019.
- [153] Oran Gafni, Adam Polyak, Oron Ashual, Shelly Sheynin, Devi Parikh, and Yaniv Taigman. Make-A-Scene: Scene-Based Text-to-Image Generation with Human Priors. arXiv:2203.13131, 2022.
- [154] Yuheng Li, Haotian Liu, Qingyang Wu, Fangzhou Mu, Jianwei Yang, Jianfeng Gao, Chunyuan Li, and Yong Jae Lee. GLIGEN: Open-Set Grounded Text-to-Image Generation. arXiv:2301.07093, 2023.
- [155] Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhang, Zhongang Qi, Ying Shan, and Xiaohu Qie. T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models. arXiv:2302.08453, 2023.
- [156] Dong Chen, Fangyun Wei, Ziyu Wan, Dongdong Chen, Jiawei Zhang, Jinjing Zhao, Sirui Zhang, Yang Yue, Zhiyang Liang, Baining Guo, Chong Luo, Jianmin Bao, Ji Li, Lei Shi, Qinhong Yang, Xiuyu Wu, Xuelu Feng, Yan Lu, Yanchen Dong, Yitong Wang, and Yunuo Chen. Lens: Rethinking Training Efficiency for Foundational Text-to-Image Models. arXiv:2605.21573, 2026.

- [157] Ruoyu Feng, Jinming Liu, Yuqi Wang, Xin Cheng, Boyuan Liu, Shanglin Li, Hanshen Zhu, Wenfeng Lin, Mingyu Guo, and Xin Jin. SeFi-Image: A Text-to-Image Foundation Model with Semantic-First Diffusion. arXiv:2606.22568, 2026.
- [158] Ming Ding, Zhuoyi Yang, Wenyi Hong, Wendi Zheng, Chang Zhou, Da Yin, Junyang Lin, Xu Zou, Zhou Shao, Hongxia Yang, and Jie Tang. CogView: Mastering Text-to-Image Generation via Transformers. arXiv:2105.13290, 2021.
- [159] Junsong Chen, Jincheng Yu, Chongjian Ge, Lewei Yao, Enze Xie, Yue Wu, Zhongdao Wang, James Kwok, Ping Luo, Huchuan Lu, and Zhenguo Li. PixArt- α : Fast Training of Diffusion Transformer for Photorealistic Text-to-Image Synthesis. arXiv:2310.00426, 2023.
- [160] Enze Xie, Junsong Chen, Junyu Chen, Han Cai, Haotian Tang, Yujun Lin, Zhekai Zhang, Muyang Li, Ligeng Zhu, Yao Lu, and Song Han. SANA: Efficient High-Resolution Image Synthesis with Linear Diffusion Transformers. arXiv:2410.10629, 2024.
- [161] Xiaokang Chen, Zhiyu Wu, Xingchao Liu, Zizheng Pan, Wen Liu, Zhenda Xie, Xingkai Yu, and Chong Ruan. Janus-Pro: Unified Multimodal Understanding and Generation with Data and Model Scaling. arXiv:2501.17811, 2025.
- [162] Tencent Hunyuan Foundation Model Team. HunyuanImage 3.0 Technical Report. arXiv:2509.23951, 2025.
- [163] Z-Image Team, Huanqia Cai, Sihan Cao, Ruoyi Du, Peng Gao, Aiming Hao, Steven Hoi, Zhaohui Hou, Shijie Huang, Dengyang Jiang, Yuming Jiang, Xin Jin, Liangchen Li, Zhen Li, Zhong-Yu Li, David Liu, Dongyang Liu, Qilong Wu, Feng Yu, Zechao Zhan, Chi Zhang, Shifeng Zhang, Ruikai Zhou, and Shilin Zhou. Z-Image: An Efficient Image Generation Foundation Model with Single-Stream Diffusion Transformer. arXiv:2511.22699, 2025.
- [164] Guillaume Le Moing, Tuan-Hung Vu, Himalaya Jain, Patrick Pérez, and Matthieu Cord. Semantic Palette: Guiding Scene Generation with Class Proportions. arXiv:2106.01629, 2021.
- [165] Hila Chefer, Yuval Alaluf, Yael Vinker, Lior Wolf, and Daniel Cohen-Or. Attend-and-Excite: Attention-Based Semantic Guidance for Text-to-Image Diffusion Models. arXiv:2301.13826, 2023.
- [166] Jinheng Xie, Yuexiang Li, Yawen Huang, Haozhe Liu, Wentian Zhang, Yefeng Zheng, and Mike Zheng Shou. BoxDiff: Text-to-Image Synthesis with Training-Free Box-Constrained Diffusion. arXiv:2307.10816, 2023.
- [167] Amir Hertz, Andrey Voynov, Shlomi Fruchter, and Daniel Cohen-Or. Style Aligned Image Generation via Shared Attention. arXiv:2312.02133, 2023.
- [168] Sicheng Mo, Fangzhou Mu, Kuan Heng Lin, Yanli Liu, Bochen Guan, Yin Li, and Bolei Zhou. FreeControl: Training-Free Spatial Control of Any Text-to-Image Diffusion Model with Any Condition. arXiv:2312.07536, 2023.
- [169] Ryugo Morita, Stanislav Frolov, Brian Bernhard Moser, Ko Watanabe, Riku Takahashi, and Andreas Dengel. LGTM: Training-Free Light-Guided Text-to-Image Diffusion Model via Initial Noise Manipulation. arXiv:2603.24086, 2026.
- [170] Linyan Dai, Xinwei Zhang, Haoyang Li, Qingqing Ye, and Haibo Hu. IDDM: Identity-Decoupled Personalized Diffusion Models with a Tunable Privacy-Utility Trade-off. arXiv:2604.00903, 2026.
- [171] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. arXiv:2208.12242, 2022.
- [172] Panagiotia Gatoula, Grigoris Karypidis, and Dimitris K. Iakovidis. EndoFSA: Endoscopic Few-Shot Image Generation via Rank-Constrained Parameter Adaptation. arXiv:2609.29930, 2026.
- [173] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training Generative Adversarial Networks with Limited Data. arXiv:2006.06676, 2020.
- [174] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. Transferring GANs: generating images from limited data. arXiv:1805.01677, 2018.
- [175] Yaxing Wang, Abel Gonzalez-Garcia, David Berga, Luis Herranz, Fahad Shahbaz Khan, and Joost van de Weijer. MineGAN: effective knowledge transfer from GANs to target domains with few images. arXiv:1912.05270, 2019.
- [176] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. Freeze the Discriminator: a Simple Baseline for Fine-Tuning GANs. arXiv:2002.10964, 2020.
- [177] Giorgio Giannone, Didrik Nielsen, and Ole Winther. Few-Shot Diffusion Models. arXiv:2205.15463, 2022.

- [178] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H. Bermano, Gal Chechik, and Daniel Cohen-Or. An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion. arXiv:2208.01618, 2022.
- [179] Ziyi Dong, Pengxu Wei, and Liang Lin. DreamArtist: Towards Controllable One-Shot Text-to-Image Generation via Contrastive Prompt-Tuning. arXiv:2211.11337, 2022.
- [180] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multi-Concept Customization of Text-to-Image Diffusion. arXiv:2212.04488, 2022.
- [181] Zhiheng Liu, Ruili Feng, Kai Zhu, Yifei Zhang, Kecheng Zheng, Yu Liu, Deli Zhao, Jingren Zhou, and Yang Cao. Cones: Concept Neurons in Diffusion Models for Customized Generation. arXiv:2303.05125, 2023.
- [182] Ligong Han, Yinxiao Li, Han Zhang, Peyman Milanfar, Dimitris Metaxas, and Feng Yang. SVDiff: Compact Parameter Space for Diffusion Fine-Tuning. arXiv:2303.11305, 2023.
- [183] Amit Raj, Srinivas Kaza, Ben Poole, Michael Niemeyer, Nataniel Ruiz, Ben Mildenhall, Shiran Zada, Kfir Aberman, Michael Rubinstein, Jonathan Barron, Yuanzhen Li, and Varun Jampani. DreamBooth3D: Subject-Driven Text-to-3D Generation. arXiv:2303.13508, 2023.
- [184] Yoad Tewel, Rinon Gal, Gal Chechik, and Yuval Atzmon. Key-Locked Rank One Editing for Text-to-Image Personalization. arXiv:2305.01644, 2023.
- [185] Hong Chen, Yipeng Zhang, Simin Wu, Xin Wang, Xuguang Duan, Yuwei Zhou, and Wenwu Zhu. DisenBooth: Identity-Preserving Disentangled Tuning for Subject-Driven Text-to-Image Generation. arXiv:2305.03374, 2023.
- [186] Yuchao Gu, Xintao Wang, Jay Zhangjie Wu, Yujun Shi, Yunpeng Chen, Zihan Fan, Wuyou Xiao, Rui Zhao, Shuning Chang, Weijia Wu, Yixiao Ge, Ying Shan, and Mike Zheng Shou. Mix-of-Show: Decentralized Low-Rank Adaptation for Multi-Concept Customization of Diffusion Models. arXiv:2305.18292, 2023.
- [187] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Wei Wei, Tingbo Hou, Yael Pritch, Neal Wadhwa, Michael Rubinstein, and Kfir Aberman. HyperDreamBooth: HyperNetworks for Fast Personalization of Text-to-Image Models. arXiv:2307.06949, 2023.
- [188] JungWoo Chae, Jiyeon Kim, JaeWoong Choi, Kyungyul Kim, and Sangheum Hwang. APT: Adaptive Personalized Training for Diffusion Models with Limited Data. arXiv:2507.02687, 2025.
- [189] Yara Bahram, Mélodie Desbos, Mohammadhadi Shateri, and Eric Granger. Uni-DAD: Unified Distillation and Adaptation of Diffusion Models for Few-step Few-shot Image Generation. arXiv:2511.18281, 2025.
- [190] Yabo Zhang, Yuxiang Wei, Dongsheng Jiang, Xiaopeng Zhang, Wangmeng Zuo, and Qi Tian. ControlVideo: Training-free Controllable Text-to-Video Generation. arXiv:2305.13077, 2023.
- [191] Harini SI, Somesh Singh, Yaman Kumar Singla, David Doermann, and Rajiv Ratn Shah. ZIPP:Zero-shot Image Personalization from Personas. arXiv:2606.08841, 2026.
- [192] Zhenhuan Wang and Fengyi Yuan. Training-Free Inpainting Across Domains with a Frozen Text-to-Image Diffusion Model. arXiv:2609.00862, 2026.
- [193] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable Augmentation for Data-Efficient GAN Training. arXiv:2006.10738, 2020.
- [194] Bingchen Liu, Yizhe Zhu, Kunpeng Song, and Ahmed Elgammal. Towards Faster and Stabilized GAN Training for High-fidelity Few-shot Image Synthesis. arXiv:2101.04775, 2021.
- [195] Tianlong Chen, Yu Cheng, Zhe Gan, Jingjing Liu, and Zhangyang Wang. Data-Efficient GAN Training Beyond (Just) Augmentations: A Lottery Ticket Perspective. arXiv:2103.00397, 2021.
- [196] Hung-Yu Tseng, Lu Jiang, Ce Liu, Ming-Hsuan Yang, and Weilong Yang. Regularizing Generative Adversarial Networks under Limited Data. arXiv:2104.03310, 2021.
- [197] Axel Sauer, Kashyap Chitta, Jens Müller, and Andreas Geiger. Projected GANs Converge Faster. arXiv:2111.01007, 2021.
- [198] Zhendong Wang, Huangjie Zheng, Pengcheng He, Weizhu Chen, and Mingyuan Zhou. Diffusion-GAN: Training GANs with Diffusion. arXiv:2206.02262, 2022.
- [199] Yao Ni and Piotr Koniusz. CHAIN: Enhancing Generalization in Data-Efficient GANs via lipsCHitz continuity constrAined Normalization. arXiv:2404.00521, 2024.
- [200] Yize Li, Yihua Zhang, Sijia Liu, and Xue Lin. Pruning then Reweighting: Towards Data-Efficient Training of Diffusion Models. arXiv:2409.19128, 2024.

- [201] Dohyun Kim, Sehwan Park, Geonhee Han, Seung Wook Kim, and Paul Hongsuck Seo. Random Conditioning with Distillation for Data-Efficient Diffusion Model Compression. arXiv:2504.02011, 2025.
- [202] Jiamu Bai, Jiaming Hu, Yanhong Wu, and Zellux Wang. Personalize at Test Time: Learning User Preferences for Image Generation. arXiv:2610.09015, 2026.
- [203] Fan Bao, Shen Nie, Kaiwen Xue, Chongxuan Li, Shi Pu, Yaole Wang, Gang Yue, Yue Cao, Hang Su, and Jun Zhu. One Transformer Fits All Distributions in Multi-Modal Diffusion at Scale. arXiv:2303.06555, 2023.
- [204] Ruiheng Zhang, Jingfeng Yao, Huangxuan Zhao, Hao Yan, Xiao He, Lei Chen, Zhou Wei, Yong Luo, Zengmao Wang, Lefei Zhang, Dacheng Tao, and Bo Du. UniX: Unifying Autoregression and Diffusion for Chest X-Ray Understanding and Generation. arXiv:2601.11522, 2026.
- [205] Jingyang Zhang, Shiwei Li, Yuanxun Lu, Tian Fang, David McKinnon, Yanghai Tsin, Long Quan, and Yao Yao. JointNet: Extending Text-to-Image Diffusion for Dense Distribution Modeling. arXiv:2310.06347, 2023.
- [206] Jiawei Mao, Yuhan Wang, Yucheng Tang, Daguang Xu, Kang Wang, Yang Yang, Zongwei Zhou, and Yuyin Zhou. MedSegFactory: Text-Guided Generation of Medical Image-Mask Pairs. arXiv:2504.06897, 2025.
- [207] Peng Wang, An Yang, Rui Men, Junyang Lin, Shuai Bai, Zhikang Li, Jianxin Ma, Chang Zhou, Jingren Zhou, and Hongxia Yang. OFA: Unifying Architectures, Tasks, and Modalities Through a Simple Sequence-to-Sequence Learning Framework. arXiv:2202.03052, 2022.
- [208] Jiasen Lu, Christopher Clark, Rowan Zellers, Roozbeh Mottaghi, and Aniruddha Kembhavi. Unified-IO: A Unified Model for Vision, Language, and Multi-Modal Tasks. arXiv:2206.08916, 2022.
- [209] Xingqian Xu, Zhangyang Wang, Eric Zhang, Kai Wang, and Humphrey Shi. Versatile Diffusion: Text, Images and Variations All in One Diffusion Model. arXiv:2211.08332, 2022.
- [210] Yen-Chi Cheng, Hsin-Ying Lee, Sergey Tulyakov, Alexander Schwing, and Liangyan Gui. SDFusion: Multi-modal 3D Shape Completion, Reconstruction, and Generation. arXiv:2212.04493, 2022.
- [211] Elad Richardson, Gal Metzer, Yuval Alaluf, Raja Giryes, and Daniel Cohen-Or. TEXTure: Text-Guided Texturing of 3D Shapes. arXiv:2302.01721, 2023.
- [212] Runpei Dong, Chunrui Han, Yuang Peng, Zekun Qi, Zheng Ge, Jinrong Yang, Liang Zhao, Jianjian Sun, Hongyu Zhou, Haoran Wei, Xiangwen Kong, Xiangyu Zhang, Kaisheng Ma, and Li Yi. DreamLLM: Synergistic Multimodal Comprehension and Creation. arXiv:2309.11499, 2023.
- [213] Jiasen Lu, Christopher Clark, Sangho Lee, Zichen Zhang, Savya Khosla, Ryan Marten, Derek Hoiem, and Aniruddha Kembhavi. Unified-IO 2: Scaling Autoregressive Multimodal Models with Vision, Language, Audio, and Action. arXiv:2312.17172, 2023.
- [214] Jinheng Xie, Weijia Mao, Zechen Bai, David Junhao Zhang, Weihao Wang, Kevin Qinghong Lin, Yuchao Gu, Zhijie Chen, Zhenheng Yang, and Mike Zheng Shou. Show-o: One Single Transformer to Unify Multimodal Understanding and Generation. arXiv:2408.12528, 2024.
- [215] Shitao Xiao, Yueze Wang, Junjie Zhou, Huaying Yuan, Xingrun Xing, Ruiran Yan, Chaofan Li, Shuting Wang, Tiejun Huang, and Zheng Liu. OmniGen: Unified Image Generation. arXiv:2409.11340, 2024.
- [216] Xi Chen, Zhifei Zhang, He Zhang, Yuqian Zhou, Soo Ye Kim, Qing Liu, Yijun Li, Jianming Zhang, Nanxuan Zhao, Yilin Wang, Hui Ding, Zhe Lin, and Hengshuang Zhao. UniReal: Universal Image Generation and Editing via Learning Real-world Dynamics. arXiv:2412.07774, 2024.
- [217] Chaorui Deng, Deyao Zhu, Kunchang Li, Chenhui Gou, Feng Li, Zeyu Wang, Shu Zhong, Weihao Yu, Xiaonan Nie, Ziang Song, Guang Shi, and Haoqi Fan. Emerging Properties in Unified Multimodal Pretraining. arXiv:2505.14683, 2025.
- [218] Fengyi Fu, Mengqi Huang, Shaojin Wu, Yunsheng Jiang, Yufei Huo, Hao Li, Yinghang Song, Fei Ding, Jianzhu Guo, Qian He, Zheren Fu, Zhendong Mao, and Yongdong Zhang. Lance: Unified Multimodal Modeling by Multi-Task Synergy. arXiv:2605.18678, 2026.
- [219] Zhipeng Bao, Zhen Zhu, Nupur Kumari, Anurag Bagchi, Yu-Xiong Wang, Pavel Tokmakov, and Martial Hebert. UniGen-AR: Unifying Visual Generation with Auto-Regressive Modeling. arXiv:2607.24157, 2026.
- [220] Junliang Ye, Kenkun Liu, Guocun Wang, Yang Li, Yansong Qu, Chunshi Wang, Jingwei Xu, Yunhan Yang, Zibo Zhao, Jiachen Xu, Jiaao Yu, Lifu Wang, Zhihao Liang, Xin Huang, Zhuo Chen, and Chunchao Guo. Hunyuan3D-Buffalo 1.0: A Unified Multimodal Model for Scalable 3D Generation, Understanding, and Editing. arXiv:2608.02711, 2026.
- [221] Zineng Tang, Ziyi Yang, Chenguang Zhu, Michael Zeng, and Mohit Bansal. Any-to-Any Generation via Composable Diffusion. arXiv:2305.11846, 2023.

- [222] Quan Sun, Qiying Yu, Yufeng Cui, Fan Zhang, Xiaosong Zhang, Yueze Wang, Hongcheng Gao, Jingjing Liu, Tiejun Huang, and Xinlong Wang. Emu: Generative Pretraining in Multimodality. *arXiv:2307.05222*, 2023.
- [223] Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. NExT-GPT: Any-to-Any Multimodal LLM. *arXiv:2309.05519*, 2023.
- [224] Zineng Tang, Ziyi Yang, Mahmoud Khademi, Yang Liu, Chenguang Zhu, and Mohit Bansal. CoDi-2: In-Context, Interleaved, and Interactive Any-to-Any Generation. *arXiv:2311.18775*, 2023.
- [225] Jun Zhan, Junqi Dai, Jiasheng Ye, Yunhua Zhou, Dong Zhang, Zhigeng Liu, Xin Zhang, Ruibin Yuan, Ge Zhang, Linyang Li, Hang Yan, Jie Fu, Tao Gui, Tianxiang Sun, Yu-Gang Jiang, and Xipeng Qiu. AnyGPT: Unified Multimodal LLM with Discrete Sequence Modeling. *arXiv:2402.12226*, 2024.
- [226] Chameleon Team. Chameleon: Mixed-Modal Early-Fusion Foundation Models. *arXiv:2405.09818*, 2024.
- [227] Chunting Zhou, Lili Yu, Arun Babu, Kushal Tirumala, Michihiro Yasunaga, Leonid Shamis, Jacob Kahn, Xuezhe Ma, Luke Zettlemoyer, and Omer Levy. Transfusion: Predict the Next Token and Diffuse Images with One Multi-Modal Model. *arXiv:2408.11039*, 2024.
- [228] Xinlong Wang, Xiaosong Zhang, Zhengxiong Luo, Quan Sun, Yufeng Cui, Jinsheng Wang, Fan Zhang, Yueze Wang, Zhen Li, Qiying Yu, Yingli Zhao, Yulong Ao, Xuebin Min, Tao Li, Boya Wu, Bo Zhao, Bowen Zhang, Liangdong Wang, Guang Liu, Zheqi He, Xi Yang, Jingjing Liu, Yonghua Lin, Tiejun Huang, and Zhongyuan Wang. Emu3: Next-Token Prediction is All You Need. *arXiv:2409.18869*, 2024.
- [229] Shufan Li, Konstantinos Kallidromitis, Akash Gokul, Zichun Liao, Yusuke Kato, Kazuki Kozuka, and Aditya Grover. OmniFlow: Any-to-Any Generation with Multi-Modal Rectified Flows. *arXiv:2412.01169*, 2024.
- [230] Kyeongmin Yeo, Yunhong Min, and Minhyuk Sung. MUNI: Multimodal Unified Latent Diffusion for Coherent Any-to-Any Generation. *arXiv:2606.16408*, 2026.
- [231] Ron Slossberg, Gil Shamaï, and Ron Kimmel. High Quality Facial Surface and Texture Synthesis via Generative Adversarial Networks. *arXiv:1808.08281*, 2018.
- [232] Yuxuan Zhang, Huan Ling, Jun Gao, Kangxue Yin, Jean-Francois Lafleche, Adela Barriuso, Antonio Torralba, and Sanja Fidler. DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort. *arXiv:2104.06490*, 2021.
- [233] Daiqing Li, Huan Ling, Seung Wook Kim, Karsten Kreis, Adela Barriuso, Sanja Fidler, and Antonio Torralba. BigDatasetGAN: Synthesizing ImageNet with Pixel-wise Annotations. *arXiv:2201.04684*, 2022.
- [234] Jun Gao, Tianchang Shen, Zian Wang, Wenzheng Chen, Kangxue Yin, Daiqing Li, Or Litany, Zan Gojcic, and Sanja Fidler. GET3D: A Generative Model of High Quality 3D Textured Shapes Learned from Images. *arXiv:2209.11163*, 2022.
- [235] Gabriela Ben Melech Stan, Diana Wofk, Scottie Fox, Alex Redden, Will Saxton, Jean Yu, Estelle Aflalo, Shao-Yen Tseng, Fabio Nonato, Matthias Muller, and Vasudev Lal. LDM3D: Latent Diffusion Model for 3D. *arXiv:2305.10853*, 2023.
- [236] Giuseppe Vecchio, Renato Sortino, Simone Palazzo, and Concetto Spampinato. MatFuse: Controllable Material Generation with Diffusion Models. *arXiv:2308.11408*, 2023.
- [237] Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, and Wenping Wang. Wonder3D: Single Image to 3D using Cross-Domain Diffusion. *arXiv:2310.15008*, 2023.
- [238] Lingteng Qiu, Guanying Chen, Xiaodong Gu, Qi Zuo, Mutian Xu, Yushuang Wu, Weihao Yuan, Zilong Dong, Liefeng Bo, and Xiaoguang Han. RichDreamer: A Generalizable Normal-Depth Diffusion Model for Detail Richness in Text-to-3D. *arXiv:2311.16918*, 2023.
- [239] Lu Qi, Lehan Yang, Weidong Guo, Yu Xu, Bo Du, Varun Jampani, and Ming-Hsuan Yang. UniGS: Unified Representation for Image Generation and Segmentation. *arXiv:2312.01985*, 2023.
- [240] Yawar Siddiqui, Tom Monnier, Filippos Kokkinos, Mahendra Kariya, Yanir Kleiman, Emilien Garreau, Oran Gafni, Natalia Neverova, Andrea Vedaldi, Roman Shapovalov, and David Novotny. Meta 3D AssetGen: Text-to-Mesh Generation with High-Quality Geometry, Texture, and PBR Materials. *arXiv:2407.02445*, 2024.
- [241] Kwon Byung-Ki, Qi Dai, Lee Hyoseok, Chong Luo, and Tae-Hyun Oh. JointDiT: Enhancing RGB-Depth Joint Modeling with Diffusion Transformers. *arXiv:2505.00482*, 2025.
- [242] Jiabin Huang, Yuanbo Yang, Bangbang Yang, Lin Ma, Yuwen Ma, and Yiyi Liao. Gen3R: 3D Scene Generation Meets Feed-Forward Reconstruction. *arXiv:2601.04090*, 2026.

- [243] Jon Hasselgren, Zheng Zeng, Milos Hasan, and Jacob Munkberg. VideoMatGen: PBR Materials through Joint Generative Modeling. arXiv:2603.16566, 2026.
- [244] Bardenus Pieter Duisterhof, Deva Ramanan, Jeffrey Ichnowski, Justin Johnson, and Keunhong Park. Modality Forcing for Scalable Spatial Generation. arXiv:2606.13676, 2026.
- [245] Xiang Chen, Yan Xia, Erica Dall'Armellina, Nishant Ravikumar, and Alejandro F Frangi. Joint shape/texture representation learning for cardiovascular disease diagnosis from magnetic resonance imaging. Primary source, 2024.
- [246] Guandao Yang, Xun Huang, Zekun Hao, Ming-Yu Liu, Serge Belongie, and Bharath Hariharan. PointFlow: 3D Point Cloud Generation with Continuous Normalizing Flows. arXiv:1906.12320, 2019.
- [247] Keyu Tian, Yi Jiang, Zehuan Yuan, Bingyue Peng, and Liwei Wang. Visual Autoregressive Modeling: Scalable Image Generation via Next-Scale Prediction. arXiv:2404.02905, 2024.
- [248] Yang Zhou, Zhen Zhu, Xiang Bai, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Non-Stationary Texture Synthesis by Adversarial Expansion. arXiv:1805.04487, 2018.
- [249] Michael Oechsle, Lars Mescheder, Michael Niemeyer, Thilo Strauss, and Andreas Geiger. Texture Fields: Learning Texture Representations in Function Space. arXiv:1905.07259, 2019.
- [250] Philipp Henzler, Niloy J. Mitra, and Tobias Ritschel. Learning a Neural 3D Texture Space from 2D Exemplars. arXiv:1912.04158, 2019.
- [251] Yu Guo, Cameron Smith, Miloš Hašan, Kalyan Sunkavalli, and Shuang Zhao. MaterialGAN: Reflectance Capture using a Generative SVBRDF Model. arXiv:2010.00114, 2020.
- [252] Yawar Siddiqui, Justus Thies, Fangchang Ma, Qi Shan, Matthias Nießner, and Angela Dai. Texturify: Generating Textures on 3D Shape Surfaces. arXiv:2204.02411, 2022.
- [253] Xilong Zhou, Miloš Hašan, Valentin Deschaintre, Paul Guerrero, Kalyan Sunkavalli, and Nima Kalantari. TileGen: Tileable, Controllable Material Generation and Capture. arXiv:2206.05649, 2022.
- [254] Tianshi Cao, Karsten Kreis, Sanja Fidler, Nicholas Sharp, and Kangxue Yin. TexFusion: Synthesizing 3D Textures with Text-Guided Image Diffusion Models. arXiv:2310.13772, 2023.
- [255] Xianfang Zeng, Xin Chen, Zhongqi Qi, Wen Liu, Zibo Zhao, Zhibin Wang, Bin Fu, Yong Liu, and Gang Yu. Paint3D: Paint Anything 3D with Lighting-Less Texture Diffusion Models. arXiv:2312.13913, 2023.
- [256] Kangle Deng, Timothy Omernick, Alexander Weiss, Deva Ramanan, Jun-Yan Zhu, Tinghui Zhou, and Maneesh Agrawala. FlashTex: Fast Relightable Mesh Texturing with LightControlNet. arXiv:2402.13251, 2024.
- [257] Yuqing Zhang, Yuan Liu, Zhiyu Xie, Lei Yang, Zhongyuan Liu, Mengzhou Yang, Runze Zhang, Qilong Kou, Cheng Lin, Wenping Wang, and Xiaogang Jin. DreamMat: High-quality PBR Material Generation with Geometry- and Light-aware Diffusion Models. arXiv:2405.17176, 2024.
- [258] Raphael Bensadoun, Yanir Kleiman, Idan Azuri, Omri Harosh, Andrea Vedaldi, Natalia Neverova, and Oran Gafni. Meta 3D TextureGen: Fast and Consistent Texture Generation for 3D Objects. arXiv:2407.02430, 2024.
- [259] Team Hunyuan3D, Shuhui Yang, Mingxin Yang, Yifei Feng, Xin Huang, Sheng Zhang, Zebin He, Di Luo, Haolin Liu, Yunfei Zhao, Qingxiang Lin, Zeqiang Lai, Xianghui Yang, Huiwen Shi, Zibo Zhao, Bowen Zhang, Hongyu Yan, Lifu Wang, Sicong Liu, Jihong Zhang, Meng Chen, Liang Dong, Yiwen Jia, Yulin Cai, Jiaao Yu, Yixuan Tang, Dongyuan Guo, Junlin Yu, Hao Zhang, Zheng Ye, Peng He, Runzhou Wu, Shida Wei, Chao Zhang, Yonghao Tan, Yifu Sun, Lin Niu, Shirui Huang, Bojian Zheng, Shu Liu, Shilin Chen, Xiang Yuan, Xiaofeng Yang, Kai Liu, Jianchen Zhu, Peng Chen, Tian Liu, Di Wang, Yuhong Liu, Linus, Jie Jiang, Jingwei Huang, and Chunchao Guo. Hunyuan3D 2.1: From Images to High-Fidelity 3D Assets with Production-Ready PBR Material. arXiv:2506.15442, 2025.
- [260] Jacob Munkberg, Peter Kocsis, and Jon Hasselgren. Texture Space Material Diffusion. arXiv:2609.37654, 2026.
- [261] Chen Yang, Jiemin Fang, Fanpeng Meng, Zihao Huang, Jichen Pang, Guanjin Wu, Zhikuan Bao, and Qi Tian. SUPRA-TEX: 3D Generation with Super-Aligned Texture. arXiv:2610.22352, 2026.
- [262] Zhiqin Chen and Hao Zhang. Learning Implicit Fields for Generative Shape Modeling. arXiv:1812.02822, 2018.
- [263] Charlie Nash, Yaroslav Ganin, S. M. Ali Eslami, and Peter W. Battaglia. PolyGen: An Autoregressive Generative Model of 3D Meshes. arXiv:2002.10880, 2020.
- [264] Ruojin Cai, Guandao Yang, Hadar Averbuch-Elor, Zekun Hao, Serge Belongie, Noah Snaveley, and Bharath Hariharan. Learning Gradient Fields for Shape Generation. arXiv:2008.06520, 2020.

- [265] Shitong Luo and Wei Hu. Diffusion Probabilistic Models for 3D Point Cloud Generation. arXiv:2103.01458, 2021.
- [266] Lingqi Zhou, Yilun Du, and Jiajun Wu. 3D Shape Generation and Completion through Point-Voxel Diffusion. arXiv:2104.03670, 2021.
- [267] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. DreamFusion: Text-to-3D using 2D Diffusion. arXiv:2209.14988, 2022.
- [268] Xiaohui Zeng, Arash Vahdat, Francis Williams, Zan Gojcic, Or Litany, Sanja Fidler, and Karsten Kreis. LION: Latent Point Diffusion Models for 3D Shape Generation. arXiv:2210.06978, 2022.
- [269] Chen-Hsuan Lin, Jun Gao, Luming Tang, Towaki Takikawa, Xiaohui Zeng, Xun Huang, Karsten Kreis, Sanja Fidler, Ming-Yu Liu, and Tsung-Yi Lin. Magic3D: High-Resolution Text-to-3D Content Creation. arXiv:2211.10440, 2022.
- [270] Alex Nichol, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, and Mark Chen. Point-E: A System for Generating 3D Point Clouds from Complex Prompts. arXiv:2212.08751, 2022.
- [271] Zhen Liu, Yao Feng, Michael J. Black, Derek Nowrouzezahrai, Liam Paull, and Weiyang Liu. MeshDiffusion: Score-based Generative 3D Mesh Modeling. arXiv:2303.08133, 2023.
- [272] Heewoo Jun and Alex Nichol. Shap-E: Generating Conditional 3D Implicit Functions. arXiv:2305.02463, 2023.
- [273] Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan Li, Hang Su, and Jun Zhu. ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation. arXiv:2305.16213, 2023.
- [274] Yawar Siddiqui, Antonio Alliegro, Alexey Artemov, Tatiana Tommasi, Daniele Sirigatti, Vladislav Rosov, Angela Dai, and Matthias Nießner. MeshGPT: Generating Triangle Meshes with Decoder-Only Transformers. arXiv:2311.15475, 2023.
- [275] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3D Latents for Scalable and Versatile 3D Generation. arXiv:2412.01506, 2024.
- [276] Yangguang Li, Zi-Xin Zou, Zexiang Liu, Dehu Wang, Yuan Liang, Zhipeng Yu, Xingchao Liu, Yuan-Chen Guo, Ding Liang, Wanli Ouyang, and Yan-Pei Cao. TripoSG: High-Fidelity 3D Shape Synthesis using Large-Scale Rectified Flow Models. arXiv:2502.06608, 2025.
- [277] Qiang Ma, Qingjie Meng, Mengyun Qiao, Paul M. Matthews, Declan P. O'Regan, and Wenjia Bai. CardiacFlow: 3D+t Four-Chamber Cardiac Shape Completion and Generation via Flow Matching. arXiv:2509.05754, 2025.
- [278] Uriel Singer, Shelly Sheynin, Adam Polyak, Oron Ashual, Iurii Makarov, Filippos Kokkinos, Naman Goyal, Andrea Vedaldi, Devi Parikh, Justin Johnson, and Yaniv Taigman. Text-To-4D Dynamic Scene Generation. arXiv:2301.11280, 2023.
- [279] Fitsum Reda, Janne Kontkanen, Eric Tabellion, Deqing Sun, Caroline Pantofaru, and Brian Curless. FILM: Frame Interpolation for Large Motion. arXiv:2202.04901, 2022.
- [280] Boah Kim and Jong Chul Ye. Diffusion Deformable Model for 4D Temporal Medical Image Generation. arXiv:2206.13295, 2022.
- [281] David Ha and Jürgen Schmidhuber. World Models. arXiv:1803.10122, 2018.
- [282] Dani Valevski, Yaniv Leviathan, Moab Arar, and Shlomi Fruchter. Diffusion Models Are Real-Time Game Engines. arXiv:2408.14837, 2024.
- [283] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable Neural Radiance Fields. arXiv:2011.12948, 2020.
- [284] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural Radiance Fields for Dynamic Scenes. arXiv:2011.13961, 2020.
- [285] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields. arXiv:2106.13228, 2021.
- [286] Ang Cao and Justin Johnson. HexPlane: A Fast Representation for Dynamic Scenes. arXiv:2301.09632, 2023.
- [287] Yanqin Jiang, Li Zhang, Jin Gao, Weimin Hu, and Yao Yao. Consistent4D: Consistent 360-Degree Dynamic Object Generation from Monocular Video. arXiv:2311.02848, 2023.
- [288] Yuyang Zhao, Zhiwen Yan, Enze Xie, Lanqing Hong, Zhenguo Li, and Gim Hee Lee. Animate124: Animating One Image to 4D Dynamic Scene. arXiv:2311.14603, 2023.

- [289] Jiawei Ren, Liang Pan, Jiaxiang Tang, Chi Zhang, Ang Cao, Gang Zeng, and Ziwei Liu. DreamGaussian4D: Generative 4D Gaussian Splatting. arXiv:2312.17142, 2023.
- [290] Yuyang Yin, DeJia Xu, Zhangyang Wang, Yao Zhao, and Yunchao Wei. 4DGen: Grounded 4D Content Generation with Spatial-temporal Consistency. arXiv:2312.17225, 2023.
- [291] Yifei Zeng, Yanqin Jiang, Siyu Zhu, Yuanxun Lu, Youtian Lin, Hao Zhu, Weiming Hu, Xun Cao, and Yao Yao. STAG4D: Spatial-Temporal Anchored Generative 4D Gaussians. arXiv:2403.14939, 2024.
- [292] Haiyu Zhang, Xinyuan Chen, Yaohui Wang, Xihui Liu, Yunhong Wang, and Yu Qiao. 4Diffusion: Multi-view Video Diffusion Model for 4D Generation. arXiv:2405.20674, 2024.
- [293] Jiawei Ren, Kevin Xie, Ashkan Mirzaei, Hanxue Liang, Xiaohui Zeng, Karsten Kreis, Ziwei Liu, Antonio Torralba, Sanja Fidler, Seung Wook Kim, and Huan Ling. L4GM: Large 4D Gaussian Reconstruction Model. arXiv:2406.10324, 2024.
- [294] Yiming Xie, Chun-Han Yao, Vikram Voleti, Huaizu Jiang, and Varun Jampani. SV4D: Dynamic 3D Content Generation with Multi-Frame and Multi-View Consistency. arXiv:2407.17470, 2024.
- [295] Rundi Wu, Ruiqi Gao, Ben Poole, Alex Trevithick, Changxi Zheng, Jonathan T. Barron, and Aleksander Holynski. CAT4D: Create Anything in 4D with Multi-View Video Diffusion Models. arXiv:2411.18613, 2024.
- [296] Chun-Han Yao, Yiming Xie, Vikram Voleti, Huaizu Jiang, and Varun Jampani. SV4D 2.0: Enhancing Spatio-Temporal Consistency in Multi-View Video Diffusion for High-Quality 4D Generation. arXiv:2503.16396, 2025.
- [297] Renlong Wu, Haoran Chen, Yuxiang Wei, Xiaowei Jin, Wangmeng Zuo, and Hui Li. 4DHumanDiff: Direct Text-to-4DGS Generation for Consistent 360-Degree Dynamic Humans. arXiv:2607.27634, 2026.
- [298] Wenbo Bao, Wei-Sheng Lai, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang. MEMC-Net: Motion Estimation and Motion Compensation Driven Neural Network for Video Interpolation and Enhancement. arXiv:1810.08768, 2018.
- [299] Wenbo Bao, Wei-Sheng Lai, Chao Ma, Xiaoyun Zhang, Zhiyong Gao, and Ming-Hsuan Yang. Depth-Aware Video Frame Interpolation. arXiv:1904.00830, 2019.
- [300] Hyeongmin Lee, Taeoh Kim, Tae young Chung, Daehyun Pak, Yuseok Ban, and Sangyoun Lee. AdaCoF: Adaptive Collaboration of Flows for Video Frame Interpolation. arXiv:1907.10244, 2019.
- [301] Simon Niklaus and Feng Liu. Softmax Splatting for Video Frame Interpolation. arXiv:2003.05534, 2020.
- [302] Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, and Shuchang Zhou. Real-Time Intermediate Flow Estimation for Video Frame Interpolation. arXiv:2011.06294, 2020.
- [303] Tarun Kalluri, Deepak Pathak, Manmohan Chandraker, and Du Tran. FLAVR: Flow-Agnostic Video Representations for Fast Frame Interpolation. arXiv:2012.08512, 2020.
- [304] Duolikun Danier, Fan Zhang, and David Bull. LDMVFI: Video Frame Interpolation with Latent Diffusion Models. arXiv:2303.09508, 2023.
- [305] Siddhant Jain, Daniel Watson, Eric Tabellion, Aleksander Holynski, Ben Poole, and Janne Kontkanen. Video Interpolation with Diffusion Models. arXiv:2404.01203, 2024.
- [306] Guangyang Wu, Xin Tao, Changlin Li, Wenyi Wang, Xiaohong Liu, and Qingqing Zheng. Perception-Oriented Video Frame Interpolation via Asymmetric Blending. arXiv:2404.06692, 2024.
- [307] Zonglin Lyu, Ming Li, Jianbo Jiao, and Chen Chen. Frame Interpolation with Consecutive Brownian Bridge Diffusion. arXiv:2405.05953, 2024.
- [308] Jinbo Xing, Hanyuan Liu, Menghan Xia, Yong Zhang, Xintao Wang, Ying Shan, and Tien-Tsin Wong. Toon-Crafter: Generative Cartoon Interpolation. arXiv:2405.17933, 2024.
- [309] Xiaojuan Wang, Boyang Zhou, Brian Curless, Ira Kemelmacher-Shlizerman, Aleksander Holynski, and Steven M. Seitz. Generative Inbetweening: Adapting Image-to-Video Models for Keyframe Interpolation. arXiv:2408.15239, 2024.
- [310] Guozhen Zhang, Yuhan Zhu, Yutao Cui, Xiaotong Zhao, Kai Ma, and Limin Wang. Motion-Aware Generative Frame Interpolation. arXiv:2501.03699, 2025.
- [311] Zihao Zhang, Haoran Chen, Haoyu Zhao, Guansong Lu, Yanwei Fu, Hang Xu, and Zuxuan Wu. EDEN: Enhanced Diffusion for High-quality Large-motion Video Frame Interpolation. arXiv:2503.15831, 2025.
- [312] Xin You, Runze Yang, Chuyan Zhang, Zhongliang Jiang, Jie Yang, and Nassir Navab. FB-Diff: Fourier Basis-guided Diffusion for Temporal Interpolation of 4D Medical Imaging. arXiv:2507.04547, 2025.

- [313] Zonglin Lyu and Chen Chen. TLB-VFI: Temporal-Aware Latent Brownian Bridge Diffusion for Video Frame Interpolation. *arXiv:2507.04984*, 2025.
- [314] Ganggui Ding, Hao Chen, and Xiaogang Xu. FC-VFI: Faithful and Consistent Video Frame Interpolation for High-FPS Slow Motion Video Generation. *arXiv:2603.04899*, 2026.
- [315] Zhiyu Xu, Wei Shang, Yushen Xu, Wenwen Li, Zhiyan Xu, Pengfei Zhu, and Dongwei Ren. Destination-Conditioned Adaptation with Semantic-Geometric Alignment for Generative Inbetweening. *arXiv:2610.13187*, 2026.
- [316] Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Learning Latent Dynamics for Planning from Pixels. *arXiv:1811.04551*, 2018.
- [317] Danijar Hafner, Timothy Lillicrap, Jimmy Ba, and Mohammad Norouzi. Dream to Control: Learning Behaviors by Latent Imagination. *arXiv:1912.01603*, 2019.
- [318] Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering Atari with Discrete World Models. *arXiv:2010.02193*, 2020.
- [319] Vincent Micheli, Eloi Alonso, and François Fleuret. Transformers are Sample-Efficient World Models. *arXiv:2209.00588*, 2022.
- [320] Danijar Hafner, Jurgis Pasukonis, Jimmy Ba, and Timothy Lillicrap. Mastering Diverse Domains through World Models. *arXiv:2301.04104*, 2023.
- [321] Anthony Hu, Lloyd Russell, Hudson Yeo, Zak Murez, George Fedoseev, Alex Kendall, Jamie Shotton, and Gianluca Corrado. GAIA-1: A Generative World Model for Autonomous Driving. *arXiv:2309.17080*, 2023.
- [322] Sherry Yang, Yilun Du, Kamyar Ghasemipour, Jonathan Tompson, Leslie Kaelbling, Dale Schuurmans, and Pieter Abbeel. Learning Interactive Real-World Simulators. *arXiv:2310.06114*, 2023.
- [323] Eloi Alonso, Adam Jelley, Vincent Micheli, Anssi Kanervisto, Amos Storkey, Tim Pearce, and François Fleuret. Diffusion for World Modeling: Visual Details Matter in Atari. *arXiv:2405.12399*, 2024.
- [324] Jialong Wu, Shaofeng Yin, Ningya Feng, Xu He, Dong Li, Jianye Hao, and Mingsheng Long. iVideoGPT: Interactive VideoGPTs are Scalable World Models. *arXiv:2405.15223*, 2024.
- [325] Jiannan Xiang, Guangyi Liu, Yi Gu, Qiyue Gao, Yuting Ning, Yuheng Zha, Zeyu Feng, Tianhua Tao, Shibo Hao, Yemin Shi, Zhengzhong Liu, Eric P. Xing, and Zhiting Hu. Pandora: Towards General World Model with Natural Language Actions and Video States. *arXiv:2406.09455*, 2024.
- [326] Shenyuan Gao, Siyuan Zhou, Yilun Du, Jun Zhang, and Chuang Gan. AdaWorld: Learning Adaptable World Models with Latent Actions. *arXiv:2503.18938*, 2025.
- [327] Lloyd Russell, Anthony Hu, Lorenzo Bertoni, George Fedoseev, Jamie Shotton, Elahe Arani, and Gianluca Corrado. GAIA-2: A Controllable Multi-View Generative World Model for Autonomous Driving. *arXiv:2503.20523*, 2025.
- [328] Yifan Zhang, Chunli Peng, Boyang Wang, Puyi Wang, Qingcheng Zhu, Fei Kang, Biao Jiang, Zedong Gao, Eric Li, Yang Liu, and Yahui Zhou. Matrix-Game: Interactive World Foundation Model. *arXiv:2506.18701*, 2025.
- [329] Danijar Hafner, Wilson Yan, and Timothy Lillicrap. Training Agents Inside of Scalable World Models. *arXiv:2509.24527*, 2025.
- [330] Zile Wang, Zexiang Liu, Jiaying Li, Kaichen Huang, Baixin Xu, Fei Kang, Mengyin An, Peiyu Wang, Biao Jiang, Yichen Wei, Yidan Xietian, Jiangbo Pei, Liang Hu, Boyi Jiang, Hua Xue, Zidong Wang, Haofeng Sun, Wei Li, Wanli Ouyang, Xianglong He, Yang Liu, Yangguang Li, and Yahui Zhou. Matrix-Game 3.0: Real-Time and Streaming Interactive World Model with Long-Horizon Memory. *arXiv:2604.08995*, 2026.
- [331] Ze Huang, Jiahui Zhang, Hairuo Liu, Chenxi Zhang, Ran Cheng, and Li Zhang. Learning Transferable Dynamics Priors from Action to World Modeling. *arXiv:2606.29501*, 2026.
- [332] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *arXiv:1406.2661*, 2014.
- [333] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel Recurrent Neural Networks. *arXiv:1601.06759*, 2016.
- [334] Danilo Jimenez Rezende and Shakir Mohamed. Variational Inference with Normalizing Flows. *arXiv:1505.05770*, 2015.
- [335] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. *arXiv:2210.02747*, 2022.

- [336] Milad Yazdani, Yasamin Medghalchi, Pooria Ashrafian, Ilker Hacihaliloglu, and Dena Shahriari. Flow Matching for Medical Image Synthesis: Bridging the Gap Between Speed and Quality. arXiv:2503.00266, 2025.
- [337] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which Training Methods for GANs do actually Converge? arXiv:1801.04406, 2018.
- [338] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-Attention Generative Adversarial Networks. arXiv:1805.08318, 2018.
- [339] Ivan Anokhin, Kirill Demochkin, Taras Khakhulin, Gleb Sterkin, Victor Lempitsky, and Denis Korzhenkov. Image Generators with Conditionally-Independent Pixel Synthesis. arXiv:2011.13775, 2020.
- [340] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-Free Generative Adversarial Networks. arXiv:2106.12423, 2021.
- [341] Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein. Efficient Geometry-aware 3D Generative Adversarial Networks. arXiv:2112.07945, 2021.
- [342] Axel Sauer, Katja Schwarz, and Andreas Geiger. StyleGAN-XL: Scaling StyleGAN to Large Diverse Datasets. arXiv:2202.00273, 2022.
- [343] Axel Sauer, Tero Karras, Samuli Laine, Andreas Geiger, and Timo Aila. StyleGAN-T: Unlocking the Power of GANs for Fast Large-Scale Text-to-Image Synthesis. arXiv:2301.09515, 2023.
- [344] Minguk Kang, Jun-Yan Zhu, Richard Zhang, Jaesik Park, Eli Shechtman, Sylvain Paris, and Taesung Park. Scaling up GANs for Text-to-Image Synthesis. arXiv:2303.05511, 2023.
- [345] Sangeek Hyun, MinKyu Lee, and Jae-Pil Heo. Cross-scale Aligned Supervision for Training GANs. arXiv:2605.26449, 2026.
- [346] Yang Song and Stefano Ermon. Generative Modeling by Estimating Gradients of the Data Distribution. arXiv:1907.05600, 2019.
- [347] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. arXiv:2010.02502, 2020.
- [348] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the Design Space of Diffusion-Based Generative Models. arXiv:2206.00364, 2022.
- [349] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps. arXiv:2206.00927, 2022.
- [350] Tero Karras, Miika Aittala, Tuomas Kynkäänniemi, Jaakko Lehtinen, Timo Aila, and Samuli Laine. Guiding a Diffusion Model with a Bad Version of Itself. arXiv:2406.02507, 2024.
- [351] Jinkyu Kim, Jinyoung Choi, and Bohyung Han. D2PO: Optimizing Diffusion Samplers via Dynamic Preference. arXiv:2607.06609, 2026.
- [352] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. arXiv:2209.03003, 2022.
- [353] Michael S. Albergo and Eric Vanden-Eijnden. Building Normalizing Flows with Stochastic Interpolants. arXiv:2209.15571, 2022.
- [354] Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguët, Yanlei Zhang, Jarrod Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. arXiv:2302.00482, 2023.
- [355] Ricky T. Q. Chen and Yaron Lipman. Flow Matching on General Geometries. arXiv:2302.03660, 2023.
- [356] Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample Flow Matching: Straightening Flows with Minibatch Couplings. arXiv:2304.14772, 2023.
- [357] Matthew Le, Apoorv Vyas, Bowen Shi, Brian Karrer, Leda Sari, Rashel Moritz, Mary Williamson, Vimal Manohar, Yossi Adi, Jay Mahadeokar, and Wei-Ning Hsu. Voicebox: Text-Guided Multilingual Universal Speech Generation at Scale. arXiv:2306.15687, 2023.
- [358] Shivam Mehta, Ruibo Tu, Jonas Beskow, Éva Székely, and Gustav Eje Henter. Matcha-TTS: A fast TTS architecture with conditional flow matching. arXiv:2309.03199, 2023.
- [359] Xingchao Liu, Xiwen Zhang, Jianzhu Ma, Jian Peng, and Qiang Liu. InstaFlow: One Step is Enough for High-Quality Diffusion-Based Text-to-Image Generation. arXiv:2309.06380, 2023.

- [360] Nanye Ma, Mark Goldstein, Michael S. Albergo, Nicholas M. Boffi, Eric Vanden-Eijnden, and Saining Xie. SiT: Exploring Flow and Diffusion-based Generative Models with Scalable Interpolant Transformers. arXiv:2401.08740, 2024.
- [361] Benjamin Kurt Miller, Ricky T. Q. Chen, Anuroop Sriram, and Brandon M Wood. FlowMM: Generating Materials with Riemannian Flow Matching. arXiv:2406.04713, 2024.
- [362] Nicholas M. Boffi, Michael S. Albergo, and Eric Vanden-Eijnden. Flow map matching with stochastic interpolants: A mathematical framework for consistency models. arXiv:2406.07507, 2024.
- [363] Itai Gat, Tal Remez, Neta Shaul, Felix Kreuk, Ricky T. Q. Chen, Gabriel Synnaeve, Yossi Adi, and Yaron Lipman. Discrete Flow Matching. arXiv:2407.15595, 2024.
- [364] Yushen Chen, Zhikang Niu, Ziyang Ma, Keqi Deng, Chunhui Wang, Jian Zhao, Kai Yu, and Xie Chen. F5-TTS: A Fairytale that Fakes Fluent and Faithful Speech with Flow Matching. arXiv:2410.06885, 2024.
- [365] Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One Step Diffusion via Shortcut Models. arXiv:2410.12557, 2024.
- [366] Zhengyang Geng, Mingyang Deng, Xingjian Bai, J. Zico Kolter, and Kaiming He. Mean Flows for One-step Generative Modeling. arXiv:2505.13447, 2025.
- [367] Sucheng Ren, Qihang Yu, Ju He, Xiaohui Shen, Alan Yuille, and Liang-Chieh Chen. Frequency-Aware Flow Matching for High-Quality Image Generation. arXiv:2604.15521, 2026.
- [368] Francesco M. Ruscio and T. Konstantin Rusch. Low-Pass Flow Matching. arXiv:2606.02177, 2026.
- [369] Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunje Choi, and Jaejun Yoo. Reliable Fidelity and Diversity Metrics for Generative Models. arXiv:2002.09797, 2020.
- [370] Dhruva Ghosh, Hanna Hajishirzi, and Ludwig Schmidt. GenEval: An Object-Focused Framework for Evaluating Text-to-Image Alignment. arXiv:2310.11513, 2023.
- [371] Shenghuan Sun, Gregory M. Goldgof, Atul Butte, and Ahmed M. Alaa. Aligning Synthetic Medical Images with Clinical Knowledge using Human Feedback. arXiv:2306.12438, 2023.
- [372] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. arXiv:1801.03924, 2018.
- [373] Mehdi S. M. Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing Generative Models via Precision and Recall. arXiv:1806.00035, 2018.
- [374] Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved Precision and Recall Metric for Assessing Generative Models. arXiv:1904.06991, 2019.
- [375] Min Jin Chong and David Forsyth. Effectively Unbiased FID and Inception Score and where to find them. arXiv:1911.07023, 2019.
- [376] Gaurav Parmar, Richard Zhang, and Jun-Yan Zhu. On Buggy Resizing Libraries and Surprising Subtleties in FID Calculation. arXiv:2104.11222, 2021.
- [377] Jaemin Cho, Abhay Zala, and Mohit Bansal. DALL-Eval: Probing the Reasoning Skills and Social Biases of Text-to-Image Generative Transformers. arXiv:2202.04053, 2022.
- [378] Yushi Hu, Benlin Liu, Jungo Kasai, Yizhong Wang, Mari Ostendorf, Ranjay Krishna, and Noah A. Smith. TIFA: Accurate and Interpretable Text-to-Image Faithfulness Evaluation with Question Answering. arXiv:2303.11897, 2023.
- [379] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation. arXiv:2304.05977, 2023.
- [380] Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-Pic: An Open Dataset of User Preferences for Text-to-Image Generation. arXiv:2305.01569, 2023.
- [381] Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human Preference Score v2: A Solid Benchmark for Evaluating Human Preferences of Text-to-Image Synthesis. arXiv:2306.09341, 2023.
- [382] Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, and Phillip Isola. DreamSim: Learning New Dimensions of Human Visual Similarity using Synthetic Data. arXiv:2306.09344, 2023.

- [383] Kaiyi Huang, Kaiyue Sun, Enze Xie, Zhenguo Li, and Xihui Liu. T2I-CompBench: A Comprehensive Benchmark for Open-world Compositional Text-to-image Generation. arXiv:2307.06350, 2023.
- [384] Sadeep Jayasumana, Srikumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, and Sanjiv Kumar. Rethinking FID: Towards a Better Evaluation Metric for Image Generation. arXiv:2401.09603, 2023.
- [385] Zhiqiu Lin, Deepak Pathak, Baiqi Li, Jiayao Li, Xide Xia, Graham Neubig, Pengchuan Zhang, and Deva Ramanan. Evaluating Text-to-Visual Generation with Image-to-Text Generation. arXiv:2404.01291, 2024.
- [386] Ossi Räisä, Boris van Breugel, and Mihaela van der Schaar. Position: All Current Generative Fidelity and Diversity Metrics are Flawed. arXiv:2505.22450, 2025.
- [387] Niantong Li, Guangzheng Hu, Weixu Qiao, Ying Ba, Qichen Hong, Shijun Shen, Jinlin Wang, Fan Zhou, Jianye Kang, Xin Shang, Ziyi He, Wei Wang, Dalin Li, Jiahao Li, Jie Zhang, Kaiyuan Gao, Kun Yan, Lihan Jiang, Ningyuan Tang, Shengming Yin, Tianhe Wu, Xiao Xu, Xiaoyue Chen, Yuxiang Chen, Yan Shu, Yanran Zhang, Yilei Chen, Yixian Xu, Zekai Zhang, Zhendong Wang, Zihao Liu, Zikai Zhou, Hongzhu Shi, Yi Wang, Bing Zhao, Hu Wei, Lin Qu, and Chenfei Wu. Qwen-Image-Bench: From Generation to Creation in Text-to-Image Evaluation. arXiv:2605.28091, 2026.
- [388] Tim Salimans and Jonathan Ho. Progressive Distillation for Fast Sampling of Diffusion Models. arXiv:2202.00512, 2022.
- [389] Luping Liu, Yi Ren, Zhijie Lin, and Zhou Zhao. Pseudo Numerical Methods for Diffusion Models on Manifolds. arXiv:2202.09778, 2022.
- [390] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast Solver for Guided Sampling of Diffusion Probabilistic Models. arXiv:2211.01095, 2022.
- [391] Hyungjin Chung, Dohoon Ryu, Michael T. McCann, Marc L. Klasky, and Jong Chul Ye. Solving 3D Inverse Problems using Pre-trained 2D Diffusion Models. arXiv:2211.10655, 2022.
- [392] Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. UniPC: A Unified Predictor-Corrector Framework for Fast Sampling of Diffusion Models. arXiv:2302.04867, 2023.
- [393] Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent Consistency Models: Synthesizing High-Resolution Images with Few-Step Inference. arXiv:2310.04378, 2023.
- [394] Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial Diffusion Distillation. arXiv:2311.17042, 2023.
- [395] Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Frédo Durand, William T. Freeman, and Taesung Park. One-step Diffusion with Distribution Matching Distillation. arXiv:2311.18828, 2023.
- [396] Xinyin Ma, Gongfan Fang, and Xinchao Wang. DeepCache: Accelerating Diffusion Models for Free. arXiv:2312.00858, 2023.
- [397] Shanchuan Lin, Anran Wang, and Xiao Yang. SDXL-Lightning: Progressive Adversarial Diffusion Distillation. arXiv:2402.13929, 2024.
- [398] Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation Alignment for Generation: Training Diffusion Transformers Is Easier Than You Think. arXiv:2410.06940, 2024.
- [399] Shanchuan Lin, Yansong Peng, Fu-Yun Wang, and Haoqi Fan. Few-Step Generation via Data-Space Iteration. arXiv:2610.12102, 2026.
- [400] Bahjat Kawar, Gregory Vaksman, and Michael Elad. SNIPS: Solving Noisy Inverse Problems Stochastically. arXiv:2105.14951, 2021.
- [401] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhong Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising Diffusion Models for Plug-and-Play Image Restoration. arXiv:2305.08995, 2023.
- [402] Hyungjin Chung, Jeongsol Kim, and Jong Chul Ye. Direct Diffusion Bridge using Data Consistency for Inverse Problems. arXiv:2305.19809, 2023.
- [403] Bowen Song, Soo Min Kwon, Zecheng Zhang, Xinyu Hu, Qing Qu, and Liyue Shen. Solving Inverse Problems with Latent Diffusion Models via Hard Data Consistency. arXiv:2307.08123, 2023.
- [404] Imraj RD Singh, Alexander Denker, Riccardo Barbano, Željko Kereta, Bangti Jin, Kris Thielemans, Peter Maass, and Simon Arridge. Score-Based Generative Models for PET Image Reconstruction. arXiv:2308.14190, 2023.
- [405] Alper Güngör, Bahri Batuhan Bilecen, and Tolga Çukur. Bayesian Conditioned Diffusion Models for Inverse Problems. arXiv:2406.09768, 2024.

- [406] Xiaodong Wang, Ping Wang, Zhangyuan Li, and Xin Yuan. Plug-and-play Diffusion Models for Image Compressive Sensing with Data Consistency Projection. *arXiv:2509.09365*, 2025.
- [407] Yuchen Jiao, Na Li, Changxiao Cai, Yuxin Chen, and Gen Li. Provable diffusion-based posterior sampling for linear inverse problems via DDIM. *arXiv:2607.19333*, 2026.
- [408] Patricia M. Johnson, Dana J. Lin, Jure Zbontar, C. Lawrence Zitnick, Anuroop Sriram, Matthew Muckley, James S. Babb, Michael P. Recht, and Florian Knoll. Deep Learning Reconstruction Enables Prospectively Accelerated Clinical Knee MRI. Primary source, 2023.
- [409] Matthew T. Martell, Nathaniel J. M. Haven, Brendyn D. Cikaluk, Brendon S. Restall, Ewan A. McAlister, Rohan Mittal, Benjamin A. Adam, Nadia Giannakopoulos, Lashan Peiris, Sveta Silverman, Jean Deschenes, Xingyu Li, and Roger J. Zemp. Deep learning-enabled realistic virtual histology with ultraviolet photoacoustic remote sensing microscopy. Primary source, 2023.
- [410] Aldo Badano, Christian G Graff, Andreu Badal, Diksha Sharma, Rongping Zeng, Frank W Samuelson, Stephen J Glick, and Kyle J Myers. Evaluation of Digital Breast Tomosynthesis as Replacement of Full-Field Digital Mammography Using an In Silico Imaging Trial. Primary source, 2018.
- [411] Hyeri Lee, Ahmet S Ayan, Sasha Beyer, Dukagjin Blakaj, Dominic DiCostanzo, John C Grecula, Joshua Palmer, Raju Raval, Raj Singh, Evan Thomas, Michael Weldon, Xiangyu Yang, Wesley Zoller, and Nilendu Gupta. Synthetic CT artifacts in MR-only brain radiotherapy: Clinical implementation experience. Primary source, 2026.
- [412] Fadeel S. Khan, Mia K. Markey, Umberto E. Villa, Alan C. Bovik, Matthew C. Fox, Brett H. Keeling, Jason S. Reichenberg, and James W. Tunnell. Context-specific image quality assessment for virtual histologic staining: checklist and guideline. Primary source, 2026.
- [413] Adityo Prakosa, Hermenegild J Arevalo, Dongdong Deng, Patrick M Boyle, Plamen P Nikolov, Hiroshi Ashikaga, Joshua JE Blauer, Elyar Ghafoori, Carolyn J Park, Robert C Blake, Frederick T Han, Rob S MacLeod, Henry R Halperin, David J Callans, Ravi Ranjan, Jonathan Chrispin, Saman Nazarian, and Natalia A Trayanova. Personalized virtual-heart technology for guiding the ablation of infarct-related ventricular tachycardia. Primary source, 2018.
- [414] U.S. Food and Drug Administration. Assessing the Credibility of Computational Modeling and Simulation in Medical Device Submissions. Primary source, 2023.
- [415] Baptiste Vasey, Myura Nagendran, Bruce Campbell, David A Clifton, Gary S Collins, Spiros Denaxas, Alistair K Denniston, Livia Faes, Bart Geerts, Mudathir Ibrahim, Xiaoxuan Liu, Bilal A Mateen, Piyush Mathur, Melissa D McCradden, Lauren Morgan, Johan Ordish, Campbell Rogers, Suchi Saria, Daniel S W Ting, Peter Watkinson, Wim Weber, Peter Wheatstone, and Peter McCulloch. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Primary source, 2022.
- [416] Riccardo Barbano, Alexander Denker, Hyungjin Chung, Tae Hoon Roh, Simon Arridge, Peter Maass, Bangti Jin, and Jong Chul Ye. Steerable Conditional Diffusion for Out-of-Distribution Adaptation in Medical Image Reconstruction. *arXiv:2308.14409*, 2023.
- [417] Seunghoi Kim, Henry F. J. Tregidgo, Matteo Figini, Chen Jin, Sarang Joshi, and Daniel C. Alexander. Tackling Hallucination from Conditional Models for Medical Image Reconstruction with DynamicDPS. *arXiv:2503.01075*, 2025.
- [418] Seunghoi Kim, Henry F. J. Tregidgo, Chen Jin, Matteo Figini, and Daniel C. Alexander. HalluGen: Synthesizing Realistic and Controllable Hallucinations for Evaluating Image Restoration. *arXiv:2512.03345*, 2025.
- [419] Soodeh Kalaie, Andrew J. Bulpitt, Alejandro F. Frangi, and Ali Gooya. A Geometric Deep Learning Framework for Generation of Virtual Left Ventricles as Graphs. Primary source, 2024.
- [420] Haoran Dou, Seppo Virtanen, Nishant Ravikumar, and Alejandro F. Frangi. A Generative Shape Compositional Framework to Synthesise Populations of Virtual Chimaeras. *arXiv:2210.01607*, 2022.
- [421] Kenneth I Aycock, Tom Battisti, Ashley Peterson, Jiang Yao, Steven Kreuzer, Claudio Capelli, Sanjay Pant, Pras Pathmanathan, David M Hoganson, Steve M Levine, and Brent A Craven. Toward trustworthy medical device in silico clinical trials: a hierarchical framework for establishing credibility and strategies for overcoming key challenges. Primary source, 2024.