Kombination der Stärken von Simulation und generativer KI zur synthetischen Bildgenerierung

Combining the strengths of simulation and generative AI for synthetic image generation

Anne Sielemann[1], Masoud Roschani[1],
Jens Ziehn[1] und Jürgen Beyerer[12]

[1] Fraunhofer IOSB, Fraunhofer Straße 1, 76131 Karlsruhe
[2] Karlsruher Institut für Technologie (KIT),
Lehrstuhl für Interaktive Echtzeitsysteme (IES), 76131 Karlsruhe

Zusammenfassung Die Bildsynthese mittels analytischer oder datengetriebener Modelle weist jeweils individuelle Vor- und Nachteile auf. Mit hybriden Ansätzen wird versucht, die Stärken beider Ansätze zu kombinieren. Im Rahmen dieser Arbeit stellen wir unsere Rendering-Pipeline der Simulationsplattform OCTAS® und insbesondere deren Integration von generativer KI vor. Die Pipeline implementiert das Prinzip des physikalisch-basierten Renderns und beinhaltet Schnittstellen zu rasterisierungsbasierten und path-tracing-basierten Rendering Engines. Ferner enthält die Pipeline ein umfangreiches Kameramodell, um die Bilder aus dem analogen in den digitalen Bildraum zu überführen. Anhand verschiedener konkreter Anwendungsfälle erläutern wir, wie die hybriden Ansätze konkret realisiert werden können und messen den entsprechenden Effekt auf die Performanz Metriken der jeweiligen Aufgabe. Des weiteren diskutiert die Arbeit, wie hybride Synthetisierungsansätze auf weitere Anwendungsfälle übertragen werden können und was dabei zu berücksichtigen ist.

Schlüsselwörter Simulation, Bildsynthese, generative KI

Abstract Image synthesis using analytical or data-driven models each has its own specific advantages as well as disadvantages.

Hybrid approaches aim to combine the strengths of both. In this work, we present our rendering pipeline of the OCTAS® simulation platform and, in particular, its integration of generative AI. The pipeline implements the principles of physically based rendering and includes interfaces to both rasterization-based and path-tracing-based rendering engines. Furthermore, the pipeline contains an extensive camera model to transfer images from the analog to the digital image space. Using several specific application scenarios, we explain how hybrid approaches can be realized in practice and measure their effect on the performance metrics of the respective task. In addition, this work discusses how hybrid synthesis approaches can be transferred to further application scenarios and what must be taken into account in doing so.

Keywords Simulation, image synthesis, generative AI

1 Einleitung

Synthetische Bildgenerierung gewinnt zunehmend an Bedeutung, insbesondere zur Skalierung und Diversifizierung von Trainingsdaten für Anwendungen des maschinellen Lernens im Bereich Computer Vision – mit dem perspektivischen Ziel des vollumfänglichen Trainierens und Testens komplexer und ggf. sicherheitskritischer Systeme wie beispielsweise automatisierte Fahrzeuge. Zwei grundlegende Synthetisierungsansätze stehen hierbei im Vordergrund: generative KI-Modelle und physikalisch basierte Simulation bzw. 3D-Rendering. Beide verfolgen das Ziel, realitätsnahe Bilddaten bereitzustellen, unterscheiden sich jedoch hinsichtlich Eigenschaften wie Realismus, Variation, Datenbedarf, Kontrollierbarkeit und Annotationstreue.

Generative KI-Modelle lernen automatisiert und datengetrieben komplexe Korrelationen und hochdimensionale Verteilungen von Bildmerkmalen. Zu ihren zentralen Vorteilen zählen eine gute Annäherung an natürliche Bildstatistiken und ein daraus resultierender hoher Realismus bei geringen manuellen Aufwänden, allerdings um den Preis einer begrenzten Kontrollierbarkeit und ungenaueren Ground-Truth-Annotationen. Zudem sind generative Modelle von der Verfügbarkeit großer, qualitativ hochwertiger Trainingsdatensätze abhängig, die die

Zieldomäne hinreichend gut abdecken müssen und reproduzieren ggf. bestehende Verzerrungen (Bias) aus diesen Daten.

Demgegenüber stehen simulationsbasierte Verfahren, die die Bildentstehung vollumfänglich modellieren: 3D-Geometrie, Materialien, Lichtquellen, Sensorcharakteristika und physikalische Interaktionen sind explizit modelliert, parametrierbar und somit kontrollierbar. Dies ermöglicht Reproduzierbarkeit, die systematische Erzeugung seltener oder gefährlicher Szenarien, die Erweiterbarkeit von Datensätzen ohne Domänenunterschied sowie die automatische Generierung exakter Annotationen (z.B. semantische Segmentierung, Tiefeninformation). Es geht jedoch mit Aufwand für die Modellierung sowie einem potenziellen systematischen Mangel an Realismus (der sogenannte *„sim-to-real-Gap"*) einher, der die Übertragbarkeit auf reale Daten einschränkt.

Hybride Ansätze bieten das Potenzial, die Stärken beider Methoden zu kombinieren sowie deren jeweilige Schwächen abzumildern. Wir stellen in dieser Arbeit vor, wie eine Kombination beider Ansätze dazu beitragen kann, eine physikalisch korrekte Darstellung von analytisch modellierbaren Parametern zu erhalten, die auf physikalische Zusammenhänge in der Bilderzeugung zurückschließen lassen, und somit auch kritische Aspekte eines physischen Realsystems abschätzbar machen – und gleichzeitig komplexe Umgebungseigenschaften über moderne KI-Verfahren kontrollierbar abgebildet werden können. So entsteht eine Bilderzeugungskette für synthetische Daten, die sowohl variantenreich und flexibel in Bezug auf komplexe Open-World-Anwendungen ist, als auch physische Konzepte konsequent und steuerbar modelliert.

2 Stand der Technik

2.1 Generative KI zur Bildsynthese

Frühe Ansätze von generativer KI zur Bildsynthese basierten überwiegend auf *Variational Autoencoders* (VAEs) [1] und *Generative Adversarial Networks* (GANs) [2]. VAEs ermöglichen die Abbildung komplexer Bilddaten in kontinuierliche latente Räume, weisen jedoch häufig Einschränkungen in der Detailtreue auf. GAN-basierte Verfahren erzielen demgegenüber erhebliche Fortschritte hinsichtlich Fotorealismus und Auflösung, sind jedoch mit Herausforderungen wie Trai-

ningsinstabilität, Moduskollaps und begrenzter Kontrollierbarkeit verbunden. Gegenwärtig dominieren *Diffusionsmodelle* [3,4] den Stand der Technik. Diese modellieren die Bildgenerierung als iterativen Entrauschungsprozess und vereinen hohe Synthesequalität mit vergleichsweise stabilen Trainingsdynamiken. Insbesondere Latente Diffusionsmodelle [5] haben durch die Verlagerung des Diffusionsprozesses in den latenten Raum eine signifikante Effizienzsteigerung ermöglicht. Die Generierung der Bilder kann beispielsweise durch die Eingabe von Referenzbildern, Text, semantischen Masken oder Skizzen gesteuert werden. Einen tiefergehenden Einblick zu den genannten und weiteren Ansätzen gibt [6].

2.2 3D-Rendering und Simulationsansätze zur Bildsynthese

Zu den Ansätzen der simulationsbasierten Bildsynthese zählen einerseits Ansätze, die direkt Computerspiel-Engines wie *Unity*[3], *Unreal*[4] oder *Godot*[5], eine 3D-Grafiksoftware wie *Blender*[6] oder die API eines Renders verwenden. Diese Ansätze verbindet, dass sie sehr flexibel genutzt und für beliebige Anwendungsfälle angepasst werden können. Erforderlich ist jedoch ein hoher Modellierungsaufwand, da Logik implementiert werden muss, um gewünschte Szenarien aufzubauen und zu variieren. Bibliotheken wie beispielsweise *BlenderProc2* [7] können im Fall von Blender dabei unterstützen, die Komplexität zu abstrahieren. Deutlich geringerer Modellierungsaufwand besteht andererseits bei der direkten Nutzung von Computerspielen zur Bildsynthese, wie beispielsweise in [8, 9]. Diese lassen sich allerdings gar nicht oder kaum an individuelle Anwendungsfälle anpassen, und die Extraktion der Ground Truth-Daten kann eine Herausforderung darstellen. Einen Kompromiss bieten Simulatoren wie CARLA [10] oder OCTAS® [11]: im Allgemeinen vereinfachen sie den Modellierungsaufwand, sind aber gleichzeitig explizit dafür ausgelegt, individuelle Anwendungsszenarien zu simulieren.

[3] `https://unity.com`
[4] `https://www.unrealengine.com/`
[5] `https://godotengine.org`
[6] `https://www.blender.org/`

2.3 Hybride Ansätze

Das Ziel hybrider Ansätze, die Simulation und generative KI miteinander verbinden, besteht in der Regel darin, die Parametrierbarkeit und damit die Kontrollierbarkeit simulationsbasierter Verfahren mit dem visuellen Realismus und der Varianz von generativer KI zu vereinen. Hierzu werden typischerweise entweder einzelne Komponenten des simulationsbasierten Syntheseprozesses durch generative KI ersetzt oder die Prozesskette um eine KI-gestützte Nachbearbeitung erweitert. Im erstgenannten Fall kann beispielsweise die Texturgenerierung für 3D-Geometrien durch generative KI substituiert werden [12]. Darüber hinaus wird intensiv daran geforscht, KI in den Renderingprozess zu integrieren. Dabei werden aus Bilddaten Szenenrepräsentationen gelernt, aus denen anschließend neue Ansichten synthetisiert werden können. Beispiele hierfür sind *Neural Radiance Fields*, kurz *NeRF* [13], oder *3D Gaussian Splatting* [14]. Zur Nachbearbeitung von Simulationsbildern durch generative KI ist es insbesondere populär, neuronalen *Style Transfer* [15] anzuwenden, um die generierten Daten der Zieldomäne stärker anzugleichen. Zudem kann die Datenbasis mittels neuronaler *Inpaintings* [16] erweitert oder einzelne Bildeigenschaften gezielt mittels sogenanntem *Attribute Editing* verändert werden [17].

3 Methodik

Auf abstraktem Level besteht die OCTAS® Rendering Pipeline zur Generierung von Sensordaten aus vier Bausteinen: (1) dem Aufbau des 3D-Szenarios im Objektraum, (2) der Überführung der Szene vom Objektraum in den (analogen) Bildraum, (3) der Simulation von Kameraeffekten im analogen Bildraum und (4) der Überführung vom analogen in den digitalen Bildraum durch Diskretisierung und Simulation von Digitalisierungseffekten. Ein schematischer Überblick mit Darstellung der Unterschritte jedes dieser Pipeline-Bausteine ist in Abbildung 1 gegeben. Das Design der Pipeline folgt insbesondere zwei leitenden Designprinzipien: Modularität und physikalisch basiertes Rendern, engl. *Physically Based Rendering* (PBR) [18]. Das Ziel von PBR ist dabei, den gesamten Renderingprozess, insbesondere die Wechselwirkungen von Licht und Materialien, mittels physikalisch plausibler Modelle so abzubilden, dass realistische, konsistente und szenenübergreifend vor-

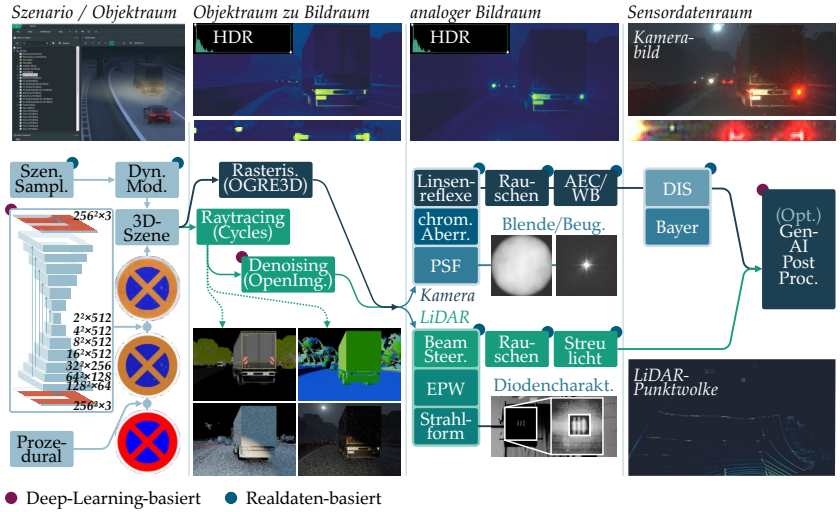


Abbildung 1: Schematische Darstelltung der Rendering Pipeline von OCTAS®.

hersagbare Bildresultate entstehen. Auf diese Weise unterstützt PBR das Designprinzip der Modularität, da sich Komponenten wie etwa die eingesetzte Rendering-Engine dank der standardisierten Darstellung von Materialien, Lichtern etc. einfach und dennoch vergleichbar austauschen lassen.

3.1 Prinzipien der Integration von generativer KI

Die beiden leitenden Designprinzipien finden auch bei der Integration von generativer KI in die OCTAS® Rendering Pipeline Anwendung. Während die Modularität es prinzipiell erlaubt, jede Komponente auszutauschen und durch ein KI-Pendant zu ersetzen, erfordert PBR eine physikalisch basierte Modellierung und Bildberechnung. Allerdings kann die physikalische Korrektheit der Beleuchtung durch generative KI nur bedingt gewährleistet werden, da Beleuchtung aus dem komplexen Zusammenspiel von Geometrie, Material, Lichtquelle und Sichtwinkel hervorgeht. Aktuelle bildgenerierende Modelle erlernen in der Regel primär statistische Korrelationen aus zweidimen-

sionalen Bilddaten, verfügen jedoch nicht über ein vollständiges physikalisches Verständnis der zugrunde liegenden 3D-Szene. Aus diesem Grund wird in der OCTAS® Rendering-Pipeline derzeit vermieden, Aufgaben wie die Beleuchtungsberechnung oder die umfassende Darstellung von Materialeigenschaften durch generative KI zu ersetzen. Ebenso werden keine ausschließlich KI-generierten Objekte mittels neuronaler *Inpaintings* nachträglich in die Bilder eingefügt.

Bei Texturen stellt sich die Situation hingegen anders dar. Diese lassen sich nur mit erheblichem Aufwand analytisch modellieren, da sie durch hohe Variabilität, Unregelmäßigkeiten und zahlreiche lokale Abweichungen gekennzeichnet sind. Sie folgen daher in der Regel keinen einfachen, geschlossenen mathematischen Gesetzmäßigkeiten, sondern entstehen aus komplexen stochastischen Prozessen. Eine manuelle Modellierung und Parametrisierung ist entsprechend aufwendig, insbesondere wenn realitätsnahe Ergebnisse erzielt werden sollen. Generative KI kann Textureigenschaften hingegen vergleichsweise zuverlässig lernen, da diese überwiegend durch lokale und statistisch wiederkehrende Bildmuster charakterisiert sind. Voraussetzung hierfür ist jedoch, dass die Trainingsdaten eine ausreichende Variationsbreite aufweisen und möglichst frei von eingebackenen Beleuchtungsartefakten sind. Nur so kann eine konsistente Verwendung der generierten Texturen unter unterschiedlichen Beleuchtungsszenarien gewährleistet werden. Darüber hinaus sollten die resultierenden Texturen möglichst PBR-tauglich sein. Das bedeutet insbesondere, dass ergänzende Texturen, beispielsweise für Normalen, Rauheit oder Metallizität, entweder gemeinsam oder nachträglich generiert werden können.

Eine Zwischenposition nimmt der *Style Transfer* ein, der sowohl zur *Domain Adaptation* als auch zur Steigerung des Realismus simulierter Bilder eingesetzt werden kann. Da sich die generative KI hierbei an den von der Simulation vorberechneten Szenenstrukturen und Beleuchtungsverhältnissen orientieren kann, bleiben wesentliche globale Strukturen, insbesondere Schattenwürfe und Lichtrichtungen, in der Regel annähernd erhalten. Dadurch sind die Ergebnisse tendenziell realistischer, als dies bei einer vollständigen Generierung vergleichbarer Bilder durch KI der Fall wäre. Jedoch können lokale Strukturen wie Objektformen und -details, Schattengrenzen, Reflexionen und Glanzlichter von der KI modifiziert werden. Zudem überträgt der Style Transfer in der Regel auch die Farbverteilung und den Kontrast, wo-



Abbildung 2: Beispiele für mit der OCTAS® Rendering Pipeline generierte synthetische Bilder zur Verkehrszeichenklassifikation mit einfachen Templates als Texturen (obere Reihe) und durch ein GAN variierte Texturen (untere Reihe).

durch die physikalische Korrektheit in der Regel nicht gewährleistet ist. Für Anwendungen, die hohe Anforderungen an die physikalische Korrektheit der Daten stellen, ist die Erzeugung von Trainingsdaten mittels Style Transfer daher nicht zu empfehlen bzw. nur mit sorgfältiger Prüfung durchzuführen. Zur Domain Adaptation kann dieses Verfahren hingegen ein geeignetes und wirkungsvolles Werkzeug darstellen.

4 Anwendungsbeispiele

4.1 Daten zur Verkehrszeichenklassifikation

Verkehrszeichen sind einer Vielzahl externer Einflüsse ausgesetzt, wie beispielsweise Sonnenlicht, Witterung, Temperaturwechsel oder Vandalismus. Diese verursachen unter anderem das Ausbleichen der Farbe, Risse in der Oberfläche, Ablagerungen von Dreck oder Verformungen, die die Oberfläche jeder einzelnen Verkehrszeicheninstanz innerhalb einer Klasse unterschiedlich aussehen lassen. Denen im vorherigen Abschnitt vorgestellten Prinzipien folgend, sollen diese Effekte berücksichtigt werden, indem die individuellen Verkehrszeichentexturen mittels generativer KI erzeugt werden. Dabei nutzen wir ein Pix2Pix GAN [19], um realistische Verkehrszeichenmerkmale auf prozedural erzeugte Templates zu übertragen, ohne deren Struktur zu verändern, wodurch der Grad der Degradierung parametrierbar bleibt (vgl. Abbildung 1). Mehr Informationen zum Generierungsprozess in [20,21]. Um den Effekt auf die Performance zu messen, vergleichen wir die Klassifikationsergebnisse von zwei *ConvNext V2* Netzwerken [22],



Abbildung 3: Beispiele des Style Transfers von SSB [25] in die Domäne von CCSV [26] mittels drei verschiedener GAN Architekturen. Quelle: [27].

die auf zwei synthetischen Datensätzen mit jeweils 198 Bildern für jede der 43 enthaltenen Klassen trainiert wurden. Einer der Datensätze (A) nutzt mittels generativer KI erzeugte Texturen, der andere Datensatz (B) verwendet unmodifizierte Originalbilder der Verkehrszeichen zur Texturierung. Bis auf diesen Unterschied wurden beide Datensätze identisch parametriert. Abbildung 2 enthält Beispielbilder beider Datensätze. Als Validierungsdatensatz wird *GTSRB* [23] verwendet, unter Nutzung des Evaluationskonfiguration aus [24]. Gemittelt über drei Trainingsläufe mit unterschiedlichem Seed erreicht Netz (B) eine Accuracy von $0,878 \pm 0,001$. Netz (A) übertrifft diese um 0,026 und erreicht eine Accuracy von $0,904 \pm 0,007$.

4.2 Daten zur Fahrzeugmarken- und Modellerkennung (VMMR)

Die Arbeit von Krohmer et al. [27] zeigt am Beispiel der feingranularen Klassifikation von Fahrzeugmarken und -modellen (VMMR), wo die Vorteile, aber auch die Herausforderungen bei einem Style Transfer von synthetischen Daten in eine andere Zieldomäne liegen. Dabei transferieren die Autoren den mit der OCTAS® Rendering Pipeline generierten *Synset Boulevard* Datensatz (SSB) [25] in die Domäne des realen *CompCars Surveillance* Datensatzes (CCSV) [26]. Beide Domänen unterscheiden sich nicht nur durch den Realismusgrad bzw. die Erzeugungsquelle, sondern auch durch einen geografischen Bias. Abbildung 3 zeigt

Beispielresultate. Durch die Veränderung von diskriminativen Merkmalen wie beispielsweise des Markenlogos erzielen Ansätze wie CycleGAN eine Verbesserung des F1-Scores um 0,122 bei der Evaluation auf CCSV [27]. Ansätzen wie dem vorgeschlagenen FCCGAN [27], die darauf abzielen, diskriminative Features zu erhalten, gelingt sogar eine Verbesserung des F1-Scores um 0,286.

5 Zusammenfassung und Ausblick

Die Kombination von physikalisch basierter Simulation und generativer KI verbindet die Stärken beider Einzelansätze und bietet somit eine vielversprechende Grundlage für realistische, variantenreiche und zugleich kontrollierbare synthetische Datenerzeugung. Die OCTAS® Rendering-Pipeline demonstriert, wie das Framework des physikalisch basierten Renderns mit generativer KI kombiniert werden kann, sodass physikalisch modellierbare Eigenschaften wie Geometrie, Materialeigenschaften, Beleuchtung und Kameraparameter kontrolliert bleiben, während komplexe Texturen und Domänenmerkmale datengestützt ergänzt werden. Die Anwendungsbeispiele zeigen, dass dadurch sowohl die Performance als auch die Übertragbarkeit synthetischer Daten verbessert werden können. Zukünftige Forschung sollte die physikalische Plausibilität der erzeugten Inhalte weiter erhöhen und zugleich effiziente Methoden zur automatisierten Qualitätsbewertung entwickeln.

Literatur

1. D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
2. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” *Advances in neural information processing systems*, vol. 27, 2014.
3. J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, “Deep Unsupervised Learning using Nonequilibrium Thermodynamics,” in *International conference on machine learning*. pmlr, 2015, pp. 2256–2265.
4. J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

5. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.
6. J. Li, C. Zhang, W. Zhu, and Y. Ren, "A Comprehensive Survey of Image Generation Models Based on Deep Learning," *Annals of Data Science*, vol. 12, no. 1, pp. 141–170, 2025.
7. M. Denninger, D. Winkelbauer, M. Sundermeyer, W. Boerdijk, M. W. Knauer, K. H. Strobl, M. Humt, and R. Triebel, "BlenderProc2: A Procedural Pipeline for Photorealistic Rendering," *Journal of Open Source Software*, vol. 8, no. 82, p. 4901, 2023.
8. M. Johnson-Roberson, C. Barto, R. Mehta *et al.*, "Driving in the Matrix: Can Virtual Worlds Replace Human-Generated Annotations for Real World Tasks?" *arXiv preprint arXiv:1610.01983*, 2016, sridhar, Sharath Nittur and Rosaen, Karl and Vasudevan, Ram.
9. M. Scucchia, P. Arranz, M. Ferrara, and D. Maltoni, "From Gaming to Research: GTA V for Synthetic Data Generation for Robotics and Navigations," in *2025 7th International Conference on Robotics and Computer Vision (ICRCV)*. IEEE, 2025, pp. 187–196.
10. A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.
11. J. Ziehn, M. Roschani, A. Sielemann *et al.*, "OCTAS," https://octas.org/, 2025.
12. A. Hu, N. Desai, H. Abu Alhaija, S. W. Kim, and M. Shugrina, "Diffusion Texture Painting," in *ACM SIGGRAPH 2024 Conference Papers*, 2024.
13. B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
14. B. Kerbl, G. Kopanas, T. Leimkühler, G. Drettakis *et al.*, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Trans. Graph.*, vol. 42, no. 4, pp. 139–1, 2023.
15. L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style," *arXiv preprint arXiv:1508.06576*, 2015.
16. W. Quan, J. Chen, Y. Liu, D.-M. Yan, and P. Wonka, "Deep Learning-Based Image and Video Inpainting: A Survey," *International Journal of Computer Vision*, vol. 132, no. 7, pp. 2367–2400, 2024.

17. X. Li, Y. Chen, Y. Zhu, S. Wang, R. Zhang, and H. Xue, "ImageNet-E: Benchmarking Neural Network Robustness via Attribute Editing," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 20 371–20 381.
18. M. Pharr, W. Jakob, and G. Humphreys, *Physically based rendering: From theory to implementation*. MIT Press, 2023.
19. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.
20. A. Sielemann, L. Loercher, M.-L. Schumacher, S. Wolf, M. Roschani, J. Ziehn, and J. Beyerer, "Synset Signset Germany: a Synthetic Dataset for German Traffic Sign Recognition," in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2024, pp. 3383–3390.
21. A. Sielemann, "The Synset Signset Rendering Pipeline – A Principle for Generating Synthetic Data for ML," in *Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*, 2026.
22. S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "Convnext v2: Co-designing and scaling convnets with masked autoencoders," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 16 133–16 142.
23. J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "The German traffic sign recognition benchmark: a multi-class classification competition," in *The 2011 international joint conference on neural networks*. IEEE, 2011, pp. 1453–1460.
24. A. Sielemann, N. Bitzer, M. Roschani, J. Ziehn, and J. Beyerer, "Sensitivity Analysis of Deep-Learning Perception Models to Sensing Parameters," in *2026 IEEE 29th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2026.
25. A. Sielemann, S. Wolf, M. Roschani, J. Ziehn, and J. Beyerer, "Synset Boulevard: A Synthetic Image Dataset for VMMR," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 9146–9153.
26. L. Yang, P. Luo *et al.*, "A large-scale car dataset for fine-grained categorization and verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, change Loy, Chen and Tang, Xiaoou.
27. E. Krohmer, S. Wolf, and J. Beyerer, "Supervised domain adaptation with disjoint label spaces for fine-grained classification," in *Proceedings of the Asian Conference on Computer Vision*, 2024, pp. 50–66.