

References

- [1] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs,” *arXiv:1711.11585 [cs]*, Aug. 2018, Accessed: Oct. 11, 2021. [Online]. Available: <http://arxiv.org/abs/1711.11585>
- [2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-Image Translation with Conditional Adversarial Networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 5967–5976. doi: [10.1109/CVPR.2017.632](https://doi.org/10.1109/CVPR.2017.632).
- [3] M. Mirza and S. Osindero, “Conditional Generative Adversarial Nets,” *arXiv:1411.1784 [cs, stat]*, Nov. 2014, Accessed: Jun. 30, 2021. [Online]. Available: <http://arxiv.org/abs/1411.1784>
- [4] A. Siarohin, O. J. Woodford, J. Ren, M. Chai, and S. Tulyakov, “Motion Representations for Articulated Animation,” *arXiv:2104.11280 [cs]*, Apr. 2021, Accessed: Sep. 20, 2021. [Online]. Available: <http://arxiv.org/abs/2104.11280>
- [5] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, “First Order Motion Model for Image Animation,” *Advances in Neural Information Processing Systems*, vol. 32, 2019, Accessed: May 31, 2021. [Online]. Available: <https://proceedings.neurips.cc/paper/2019/hash/31c0b36aef265d9221af80872ceb62f9-Abstract.html>
- [6] T.-C. Wang, M.-Y. Liu, A. Tao, G. Liu, J. Kautz, and B. Catanzaro, “Few-shot Video-to-Video Synthesis,” *arXiv:1910.12713 [cs]*, Oct. 2019, Accessed: Jan. 01, 2022. [Online]. Available: <http://arxiv.org/abs/1910.12713>
- [7] O. Gafni, O. Ashual, and L. Wolf, “Single-Shot Freestyle Dance Reenactment,” *arXiv:2012.01158 [cs]*, Mar. 2021, Accessed: Dec. 27, 2021. [Online].
- [8] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros, “Everybody Dance Now,” *arXiv:1808.07371 [cs]*, Aug. 2019, Accessed: Jun. 07, 2021. [Online].
- [9] Y. Ren, G. Li, S. Liu, and T. H. Li, “Deep Spatial Transformation for Pose-Guided Person Image Generation and Animation,” *IEEE Trans. on Image Process.*, vol. 29, pp. 8622–8635, 2020,

- [10] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. Van Gool, “Pose Guided Person Image Generation,” *arXiv:1705.09368 [cs]*, Jan. 2018, Accessed: May 13, 2021. [Online].
- [11] H. Tang, S. Bai, L. Zhang, P. H. S. Torr, and N. Sebe, “XingGAN for Person Image Generation,” in *Computer Vision – ECCV 2020*, Cham, 2020, pp. 717–734.
- [12] T. Ma, B. Peng, W. Wang, and J. Dong, “MUST-GAN: Multi-Level Statistics Transfer for Self-Driven Person Image Generation,” 2021, pp. 13622–13631. Accessed: Jul. 26, 2021. [Online].
- [13] P. Zablotzkaia, A. Siarohin, B. Zhao, and L. Sigal, “DwNet: Dense warp-based network for pose-guided human video generation,” *arXiv:1910.09139 [cs]*, Oct. 2019, Accessed: Jul. 26, 2021. [Online].
- [14] I. J. Goodfellow *et al.*, “Generative Adversarial Networks,” *arXiv:1406.2661 [cs, stat]*, Jun. 2014, Accessed: Feb. 20, 2022. [Online].
- [15] J. Gui, Z. Sun, Y. Wen, D. Tao, and J. Ye, “A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications,” *arXiv:2001.06937 [cs, stat]*, Jan. 2020, Accessed: Feb. 13, 2022. [Online].
- [16] G. E. Hinton, T. J. Sejnowski, and D. H. Ackley, Boltzmann machines: Constraint satisfaction networks that learn. Carnegie-Mellon University, Department of Computer Science Pittsburgh, 1984
- [17] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” arXiv preprint arXiv:1312.6114, 2013.
- [18] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *arXiv:1505.04597 [cs]*, May 2015, Accessed: Oct. 11, 2021. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [19] R. A. Güler, N. Neverova, and I. Kokkinos, “DensePose: Dense Human Pose Estimation In The Wild,” *arXiv:1802.00434 [cs]*, Feb. 2018, Accessed: Jan. 07, 2022. [Online].
- [20] P. Li, Y. Xu, Y. Wei, and Y. Yang, “Self-Correction for Human Parsing,” *arXiv:1910.09777 [cs, eess]*, Oct. 2019, Accessed: Jan. 07, 2022. [Online].

- [21] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” *arXiv:1611.08050 [cs]*, Apr. 2017, Accessed: Jan. 07, 2022. [Online].
- [22] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv:1409.1556 [cs]*, Apr. 2015, Accessed: Mar. 11, 2022. [Online].
- [23] “Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach - ScienceDirect.” <https://www.sciencedirect.com/science/article/pii/S1877050918308019> (accessed Mar. 11, 2022).
- [24] K. N. Raju and K. S. P. Reddy, “Comparative study of Structural Similarity Index (SSIM) by using different edge detection approaches on live video frames for different color models,” in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Jul. 2017, pp. 932–937. doi: [10.1109/ICICICT1.2017.8342690](https://doi.org/10.1109/ICICICT1.2017.8342690).
- [25] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric,” *arXiv:1801.03924 [cs]*, Apr. 2018, Accessed: Mar. 11, 2022. [Online].
- [26] L. Liu *et al.*, “Neural Human Video Rendering by Learning Dynamic Textures and Rendering-to-Video Translation,” *arXiv:2001.04947 [cs]*, Jul. 2021, Accessed: Dec. 27, 2021. [Online].
- [27] T.-C. Wang *et al.*, “Video-to-Video Synthesis,” *arXiv:1808.06601 [cs]*, Dec. 2018, Accessed: Jan. 05, 2022. [Online].
- [28] S. Czolbe, O. Krause, and A. Feragen, “Semantic similarity metrics for learned image registration,” Apr. 2021, doi: [10.48550/arXiv.2104.10051](https://doi.org/10.48550/arXiv.2104.10051).