Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.1, June : 2024

AN OVERVIEW OF DEEP NEURAL NETWORKS IN AIGENERATED PAINTING, FROM PAINTBRUSH TO PIXEL

Arpitpethe, MTech Scholar, Department. of AIML, TITE , Bhopal, INDIA.
Arppethe0111@gmail.com
Shamailakhan, Professor, Department ofAIML,TITE, Bhopal, INDIA

ABSTRACT
This study discusses the several deep neural network designs and models that have been used to make art in the exciting field of AI-generated art. We look at the major players in the field, from the traditional convolutional neural networks to the state-of-the-art diffusion models. We describe these neural networks' broad architectures and underlying ideas. Next, we present milestone examples, ranging from the surreal landscapes of DeepDream to the most recent innovations, such as the captivating pictures produced by Stable Di_usion and DALL-E 2. These models are thoroughly compared, showing the advantages and disadvantages of each. Consequently, we analyze the impressive advancements that deep neural networks have made in a brief amount of time. With a distinct blend of tech-nical explanations and insights into the current state of AI-generatedart, this paper exempli_es how art and computer science interact.
KEYWORDS: Neural networks, Transformers, Di_usion models, Generativeadversarial networks, Deep learning, Image processing

INTRODUCTION
AI-generated art is a new and emerging _eld, and it will likely take some timefor it to fully mature and for its place in the art world to be fully understood,"replied ChatGPT, the chatbot created by OpenAI, when we asked about the

1.ARTICLE TITLE
state of AI-generated art at the moment. As with any new field, it will require time to gain complete understanding and establish a position for itself in the creative world.AI-generated images are commonplace today, whether or not they are acknowledged as works of art. It is no longer possible to deny their presence in our lives, regardless of debates over their level of creativity or artistic ability.

Fig. 1 (Left) Edmond de Belamy" - The _rst AI-generated portrait sold at Christie's artauction in 2018. (Right) The^atreD'opera Spatial"

Deep learning, a branch of machine learning, is responsible for all of this advancement in AI-generated art. Deep neural networks, a subfield of this one, have produced notable advances in computer vision and other fields during the past ten years. In this study, we primarily address the usage of deep neural networks for image processing and the latest advancements in this field to

create artificial intelligence-generated images. Numerous studies that approach AI-generated art from various angles may be found in the literature. For instance, Citinicet al. (2022) [7] discusses authorship, copyright, and ethical concerns in addition to the inventiveness of AI systems. In an experiment, Ragot et al. (2020) [54] asked participants to rate the difference.
in terms of preference, perceived beauty, novelty, and significance between paintings made by humans and artificial intelligence. These days, attribution of authorship and accountability for AI-generated art are also significant challenges [20]. A overview of generative AI models that covers a range of applications, including text, photos, videos, and audios, was published recently [5]. The primary neural networks that have been employed to produce realistic visuals are the subject of this research. To help readers better comprehend these models, we describe the fundamental components of the relevant neural networks. We outline the broad principles of operation of these neural networks and present the latest developments in AI-generated art, including DALL-E 2 . We look at thefast advancement of deep neural networks in artificial intelligence-generated art, and highlight key turning points in this evolution. This evaluation compares the most recent state-of-the-art models and tackles the subject from a technological standpoint. But even for a non-technical readership (such as those from more established fields in aesthetics, cultural studies, and the fine arts), this study might be useful in giving a general overview of the many methods and resources that are already accessible. This is how the remainder of the paper is structured. We go over the key neural networks that are utilized as models for AI-generated art in Section 2. In Section 3, we provide an overview of the developments and current trends in generative modeling. In Section 4, we talk about the models that are in use today. In Section 5, we wrap up.

PRELIMINARIES

A neural network modifies its weights, or key parameters, throughout training. The neural network learns when the training is over and the weights are optimum for the given job. Multilayer perceptrons (MLPs), which are helpful for classification and regression applications, are a common type of neural network. But there are a lot of deep neural networks that function very well for processing images. Convolutional neural networks (CNNs) are one of them. In this part, we begin with CNNs that learn in a supervised learning environment and require data labels during training. Next, we describe autoencoders that are capable of unsupervised learning—that is, learning without the need for data labels. GANs and the Transformer neural network are the next topics we cover. Finally, we describe the di_usion models, the most recent developments in deep learning.

CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks, sometimes called CNNs or Con-v Nets, are deep neural networks that are mostly used for image processing . Through a succession of hidden layers, ever more abstract characteristics are collected from the picture in a deep CNN. Because of this, a CNN's structure is comparable to that of the human visual cortex, with lower regions extracting basic visual elements and higher areas extracting combinations and high-level information. This splits the intricate mapping of the input pixels into a number of layered mappings., It is represented as a distinct layer for each . The literature has a variety of CNN architectures that can recognize handwritten digits, including LeNet . Deeper CNNs like AlexNet [36], VGG ,ResNet , DenseNet , _cientNet , Inception, and GoogLeNet are necessary for more complicated tasks like object recognition. Three different types of layers are often seen in a CNN: convolutional layers, poolin

2.ARTICLE TITLE

layers as well as completely linked layers. Figure 2 shows the overall architecture of a typical CNN for classification. The _rst convolutional layer comes after the input layer, where the input picture X is shown. The weights in a convolutional layer are stored inside kernels. Convolution is a mathematical procedure that takes place between the input and the kernel during learning. Convolution is essentially accomplished by multiplying each input element by each kernel element and then adding together the results of this multiplication. This input might be the output of a

convolutional layer that came before it or a picture. The convolutional molecule's unitslayer are organized into planes, called feature maps, which are defined as three-dimensional tensors. A feature map's units each get input from a little portion of the picture, and they all identify the same pattern but at different points in the input image. Since all of the units in a feature map have the same weights, a kernel convolution is used to convert the image's pixel intensities [3]. A convolutional layer often has many feature maps, each with a different weight, to detect numerous characteristics [3]. As a result, several concurrently functioning kernels make up a single convolutional layer. Furthermore, a CNN usually has millions of convolutional layers as it is a deep neural network. of parameters.

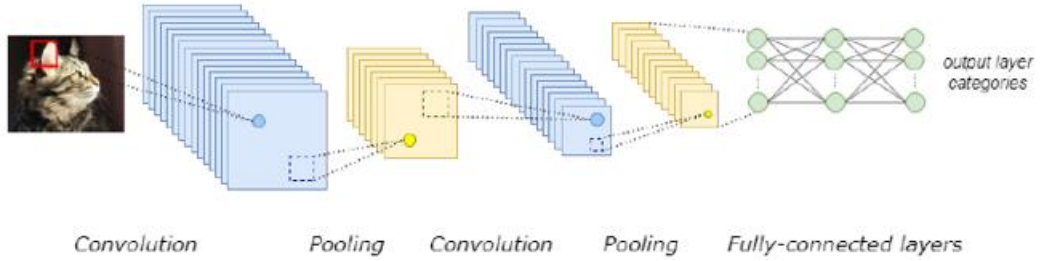

Fig. 2 An example CNN structure with two convolutional, two pooling, and three fully connected layers for classi_cation.

The recti_ed linear unit (ReLU), for example, is a nonlinear activation function that is typically applied to each output of a convolution process . Next, a pooling layer—which decreases the size of feature maps by computing a summary statistic of the nearest outputs, such as the maximum value or the average—comes after a convolutional layer or stack of convolutional layers [27]. Last but not least, fully-connected (dense) layers come after a sequence of convolutional and pooling layers. The job determines which activation function the output layer uses. For binary classification, a sigmoid function is typically favored, and softmax nonlinearity for For regression problems, multi-classification and linear activation are used. Thus, for binary and multi-class classification, respectively, a CNN minimizes the difference between the expected values ^y and the desired output values y in the cross-entropy loss functions provided in Eqs. 1 and 2. When dealing with a regression job, a CNN might reduce the mean squared error as stated in Equation 3. The weights associated with the hidden layers are denoted by W in the equations below, whereas the weights of the output layer are represented by V. While the above listed activation and loss functions are the most commonly used ones in the literature, there are additional options as well, such as as Leaky ReLU [41]as an activation function, or mean absolute error as a loss function.

$$L(W, v \mid \mathcal{X}) = -\sum_{t} y^t log\hat{y}^t + (1 - y^t) log(1 - \hat{y}^t) \quad (1)$$

$$L(W, V \mid \mathcal{X}) = -\sum_{t}\sum_{i} y_i^t log\hat{y}_i^t \quad (2)$$

$$L(W, V \mid \mathcal{X}) = \frac{1}{2}\sum_{t}\sum_{i}(y_i^t - \hat{y}_i^t)^2 \quad (3)$$

The difference between a convolutional layer and a fully-connected layer is significant because the former learns local patterns in the input feature space, while the latter learns global patterns [10]. Because of the local receptive fields, a CNN has less weights than it would have if it were totally linked [3]. CNNs have shown significant improvements for a range of image processing applications and are now considered important neural networks in deep learning.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.1, June : 2024

3. ARTICLE TITLE

A VAE is used to translate the output of the di_usion process back to pixel space. Stable Di_usion operates in (lower-dimensional) latent space inside pixel-space . This latent distortion model uses far less computer power than earlier text-to-image models, which require hundreds of GPU computing days. As a result, it is more accessible to those with limited resources. In addition to generating new pictures, Stable Di_usion enables users to alter pre-existing ones by inpainting (which involves deleting or adding elements from an existing image) or image-to-image translation (which converts a sketch into digital art).

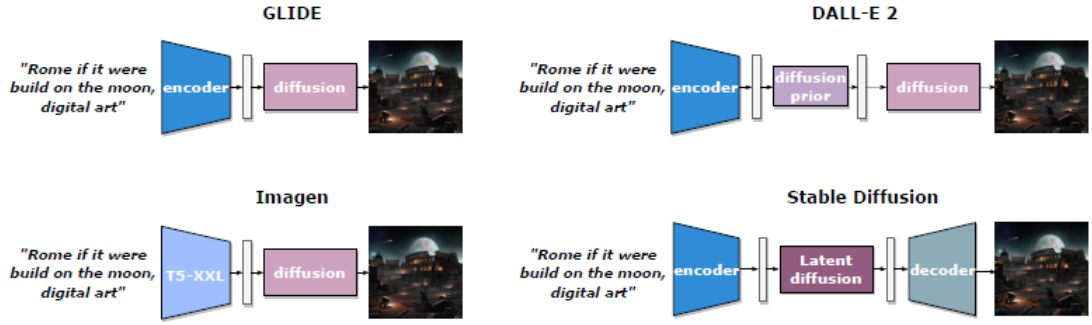

Fig. 3 Comparison of the di_usion model-based text-to-image models.

Use InstructPix2Pix. Despite the fact that a number of text-to-image models contain this inpainting functionality, it might be challenging to get the appropriate result when using a text description alone. Either the whole output picture must be described, or the user must create a mask for the intended modification and describe it for that portion of the image. A modification of stable illusion, InstructPix2Pix lets users edit photos with simple written instructions . The user may simply provide an easy command on how to alter the input picture (mask-free editing), eliminating the need to provide a mask or explain the intended output image. For example, if you would like to turn theSunowersby Vincent Van Gogh into a painting of roses, you can just writethe instruction \Swap sunowers with roses" .

Fig. 4 InstructPix2Pix turns the Sunowersby Vincent Van Gogh into a painting of roses

COMPARISON OF DEEP GENERATIVE MODELS

The vast amount of work being done in the field of generative AI makes it difficult to determine which model would be best for a given situation. Table 1 presents a comparative analysis of the aforementioned models for their computational efficiency (dataset size and trainable parameters), picture quality (FID score), model capabilities, and accessibility (open source versus proprietary).

Google's Parti, Muse, and Ima-gen are the cutting-edge models that produce the most photo-realistic pictures. Nevertheless, none of these models are publicly available, which means that only Google Research has access to them. Consequently, No other studies have been able to replicate these findings. Furthermore, these models cannot be used by artists to produce AI-generated art. Conversely, GLIDE and DALL-E 2 have filtered versions made available by OpenAI. These versions were trained using a filtered dataset that did not include any photos of violence, nudity, or identifiable individuals (political figures, celebrities, etc.). The user-friendly interface of DALL-E 2 allows the user to input a text prompt and see four produced graphics. To prevent abuse, GLIDE may be accessed via a Google Colab notebook that loads the model and its weights but conceals the origin code.

The state of AI-generated art in 2024 continues to evolve, with new developments and discussions around its impact on the art world. Here's a concise table summarizing the current landscape:

Statistic	Data
Americans who've seen AI art	27%[1]
Artists considering AI art unethical	74%[1]
Highest sale of traditional AI art	$432,000[1]
Highest sale of AI-generated NFTs	$1.1 million[1]
Ability to recognize AI art	54%[1]
Artists' income concerns	55%[1]
Artists using AI for ideas	65%[1]
Popular AI art generators	DALL-E 2, Midjourney[1]

Table 1

AI art generators like DALL-E 2 and Midjourney are at the forefront, offering a range of styles and formats. The technology behind these generators is based on machine learning models that can interpret text prompts and create digital visuals accordingly. The market for AI-generated art, including NFTs, is growing, with significant sales being made at prestigious auction houses.

However, ethical concerns and debates about the legitimacy of AI art persist. A significant portion of artists worry about the potential impact on their livelihoods, while others embrace the technology for creative inspiration. The public's ability to distinguish AI-generated art from human-created art is improving, but there's still a debate on whether AI-generated art should be considered "art" in the traditional sense.As AI art generators become more sophisticated, they're likely to continue shaping the art industry in new and unexpected ways.

CONCLUSIONS

Compared to a few years ago, deep learning and its applications in image processing are currently at a completely different level. That natural picture classification could be achieved by deep neural networks at the start of the previous decade was revolutionary.These days, these models can create intricate and incredibly realistic graphics from just text instructions. This makes it possible for those without programming experience to use these effective models. It's crucial to keep in mind that moral and responsible considerations should drive the use of these models.

REFERENCES

[1] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarialnetworks, International Conference on Machine Learning (2017) 214{223.
[2] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointlylearning to align and translate, International Conference on LearningRepresentations (2015).
[3] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[4] H. Bourlard, Y. Kamp, Auto-association by multilayer perceptrons andsingular value decomposition, Biological Cybernetics, 59 (1988) 291-294.
[5] R. Gozalo-Brizuela, E.C. Garrido-Merchan, ChatGPT is not all you need.A state of the art review of large generative AI models, arXiv preprintarXiv:2301.04655 (2023).
[6] T. Brooks, A. Holynski, A.A. Efros, InstructPix2Pix: Learning to followimage editing instructions, arXiv preprint arXiv: 2211.09800 (2022).
[7] E. Cetinic, J. She, Understanding and creating art with AI: Review andoutlook, ACM Transactions on Multimedia Computing, Communications,and Applications (TOMM) 18(2) (2022) 1{22.
[8] H. Chang et al., Muse: Text-to-image generation via masked generativetransformers, arXiv preprint arXiv:2301.00704 (2023).
[9] M. Chen et al., Generative pretraining from pixels, International Confer-ence on Machine Learning (2020) 1691{1703.
[10] F. Chollet, Deep Learning with Python, Manning Publications, 2018.
[11] J-B. Cordonnier, A. Loukas, M. Jaggi, On the relationship between self-attention and convolutional layers, International Conference on LearningRepresentations (2020).
[12] G.W. Cottrell, P. Munro, D. Zipser, Learning internal representationsfrom gray-scale images: An example of extensional programming, Proceed-ings Ninth Annual Conference of the Cognitive Science Society, (1987)
461{473.
[13] A. Creswell et al., Generative adversarial networks: An overview, IEEESignal Processing Magazine, 35(1) (2018) 53{65.
[14] J. Deng et al., ImageNet: A large-scale hierarchical image database, IEEEConference on Computer Vision and Pattern Recognition (2009) 248{255.
[15] P. Dhariwal, A. Nichol, Di_usion models beat GANs on image synthesis,Advances in Neural Information Processing Systems 34 (2021) 8780-8794.
[16] J. Sohl-Dickstein et al., Deep unsupervised learning using nonequilibriumthermodynamics, International Conference on Machine Learning (2015)2256{2265.
[17] M. Ding et al., CogView: Mastering text-to-image generation via trans-formers, Advances in Neural Information Processing Systems 34 (2021)19822{19835.
[18] A. Dosovitskiy et al., An image is worth 16 _ 16 words: Transform-ers for image recognition at scale, International Conference on LearningRepresentations (2021).
[19] A. Elgammal, B. Liu, M. Elhoseiny, M. Mazzone, CAN: Creative adver-sarial networks generating \art" by learning about styles and deviatingfrom style norms, arXiv preprint arXiv:1706.07068 (2017).

[20] Z. Epstein, S. Levine, D.G. Rand, I. Rahwan, Who gets credit for AI-generated art?,iScience 23(9) (2020) 101515.
[21] O. Gafni et al., Make-A-Scene: Scene-based text-to-image generation withhuman priors, Computer Vision{ECCV 2022: 17th European Conference(2022) 89{106.
[22] L.A. Gatys, A.S. Ecker, M. Bethge, A neural algorithm of artistic style,arXiv preprint arXiv:1508.06576 (2015).
[23] X. Glorot, A. Bordes, Y. Bengio, Deep sparse recti_er neural networks,Proceedings of the Fourteenth International Conference on Arti_cial Intel-ligence and Statistics. JMLRWorkshop and Conference Proceedings (2011)315{323.
[24] A. Graves, Supervised Sequence Labelling with Recurrent Neural Net-works, Studies in Computational Intelligence, Springer, 2012.
[25] L. Goetschalckx, A. Andonian, J. Wagemans, Generative adversarial net-works unlock new methods for cognitive science, Trends in CognitiveSciences 25(9) (2021) 788-801.
[26] I. Goodfellow et al., Generative adversarial networks, Advances in NeuralInformation Processing Systems, 27 (2014).
[27] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, The MIT Press,2016.
[28] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recogni-tion, Proceedings of the IEEE Conference on Computer Vision and PatternRecognition (2016) 770{778.
[29] J. Ho, A. Jain, P. Abbeel, Denoising di_usion probabilistic models,Advances in Neural Information Processing Systems 33 (2020) 6840{6851.
[30] G. Huang, Z. Liu, L. Maaten, K. Weinberger, Densely connected convolu-tional networks, Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition (2017) 4700{4708.
[31] P. Isola, J-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation withconditional adversarial networks, Proceedings of the IEEE Conference onComputer Vision and Pattern Recognition (2017), 1125{1134.
[32] T. Karras, S. Laine, T. Aila, A style-based generator architecture forgenerative adversarial networks, Proceedings of the IEEE/CVF Conferenceon Computer Vision and Pattern Recognition, (2019) 4401{4410.
[33] T. Karras et al., Analyzing and improving the image quality of Style-GAN, Proceedings of the IEEE/CVF Conference on Computer Vision andPattern Recognition, (2020) 8110{8119.
[34] S. Khan et al., Transformers in vision: A survey, ACM Computing Surveys54(10s) (2022) 1{41.
[35] D.P. Kingma, M. Welling, Auto-encoding variational bayes, arXivpreprint arXiv:1312.6114 (2013).
[36] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classi_cation withdeep convolutional neural networks, Advances in Neural InformationProcessing Systems 25 (2012) 1097{1105.
[37] Y. LeCun et al., Backpropagation applied to handwritten zip coderecognition, Neural Computation (1989) 541{551.
[38] Y. LeCun, L. Bottou, Y. Bengio, P. Ha_ner, Gradient-based learningapplied to document recognition, Proceedings of the IEEE 86(11) (1998)2278{2324.
[39] T-Y. Lin et al., Microsoft COCO: Common objects in context, EuropeanConference on Computer Vision (2014) 740{755.
[40] T. Lin, Y. Wang, X. Liu, X. Qui, A survey of transformers, AI Open(2022) 111{132.
[41] A.L. Maas, A.Y. Hannun, A.Y. Ng, Recti_er nonlinearities improve neuralnetwork acoustic models, International Conference on Machine Learning(2013).
[42] A. Makhzani et al., Adversarial autoencoders, arXiv preprintarXiv:1511.05644 (2015).
[43] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXivpreprint arXiv:1411.1784 (2014).
[44] A. Mordvintsev, C. Olah, M. Tyka, Inceptionism: Going deeper intoneural networks, Google Research Blog, June 18, 2015.

[45] A. Mordvintsev, C. Olah, M. Tyka, DeepDream - a code example forvisualizing neural networks, Google Research Blog, July 1, 2015.
[46] A. Nichol et al., GLIDE: Towards photorealistic image generation andediting with text-guided di_usion models, arXiv preprint arXiv:2112.10741(2021).
[47] D. Noton, L. Stark, Eye movements and visual perception, Scienti_cAmerican, 224(6) (1971) 34{43.
[48] D. Noton, L. Stark, Scanpaths in saccadic eye movements while viewingand recognizing patterns, Vision Research, 11(9) (1971) 929{942.
[49] T. Park, M-Y. Liu, T-C. Wang, J-Y. Zhu, Semantic image synthesis withspatially-adaptive normalization, Proceedings of the IEEE/CVF Confer-ence on Computer Vision and Pattern Recognition (2019) 2337{2346.
[50] N. Parmar et al., Image transformer, International Conference on MachineLearning (2018) 4055{4064.
[51] A. Radford, L. Metz, S. Chintala, Unsupervised representation learningwith deep convolutional generative adversarial networks, arXiv preprintarXiv:1511.06434 (2015).
[52] A. Radford et al., Learning transferable visual models from natural lan-guage supervision, International Conference on Machine Learning (2021)8748{8763.
[53] C. Ra_el et al., Exploring the limits of transfer learning with a uni_edtext-to-text transformer, The Journal of Machine Learning Research 21(1)(2020) 5485{5551.
[54] M. Ragot, N. Martin, S. Cojean, AI-generated vs. Human artworks. Aperception bias towards arti_cial intelligence, Extended abstracts of the2020 CHI Conference on Human Factors in Computing Systems (2020)
1{10.
[55] A. Ramesh et al., Zero-shot text-to-image generation, InternationalConference on Machine Learning (2021) 8821{8831.
[56] A. Ramesh et al., Hierarchical text-conditional image generation withCLIP latents, arXiv preprint arXiv:2204.06125 (2022).
[57] D.J. Rezende, S. Mohamed, D. Wierstra, Stochastic backpropagation andapproximate inference in deep generative models, International Conferenceon Machine Learning (2014) 1278{1286.
[58] J.T. Rolfe, Discrete variational autoencoders, International Conference onLearning Representations 2017).
[59] R. Rombach et al., High-resolution image synthesis with latent di_usionmodels, Proceedings of the IEEE/CVF Conference on Computer Visionand Pattern Recognition (2022) 10684{10695.
[60] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networksfor biomedical image segmentation, International Conference on MedicalImage Computing and Computer-Assisted Intervention (2015) 234{241.
[61] C. Saharia et al., Photorealistic text-to-image di_usion models with deeplanguage understanding, arXiv preprint arXiv:2205.11487 (2022).
[62] I. Salian, \Paint me a picture": NVIDIA research shows GauGAN AI artdemo now responds to words, NVIDIA Blog, November 22, 2021.
[63] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, International Conference on Learning Representa-tions (2015).
[64] D. Soydaner, Hyper autoencoders, Neural Processing Letters 52(2) (2020)1395{1413.
[65] D. Soydaner, Attention mechanism in neural networks: Where it comesand where it goes, Neural Computing and Applications 34 (2022) 13371{13385.
[66] C. Szegedy et al., Going deeper with convolutions, Proceedings of theIEEE Conference on Computer Vision and Pattern Recognition (2015) 1{9.
[67] W.R. Tan, C.S. Chan, H.E. Aguirre, K. Tanaka, ArtGAN: Artwork syn-thesis with conditional categorical GANs, IEEE International Conferenceon Image Processing (ICIP) (2017) 3760{3764.

[68] M. Tan, Q.V. Le, E_cientNet: Rethinking model scaling for convolutionalneural networks, International Conference on Machine Learning (2019)6105{6114.
[69] A. Vaswani et al., Attention is all you need, Advances in NeuralInformation Processing Systems 30 (2017).
[70] L. Yang et al., Di_usion models: A comprehensive survey of methods andapplications, arXiv preprint arXiv:2209.00796 (2022).
[71] J. Yu et al., Scaling autoregressive models for content-rich text-to-imagegeneration, arXiv preprint arXiv:2206.10789 (2022).
[72] J. Yu et al., Vector-quantized image modeling with improved VQGAN,International Conference on Learning Representations (2022).
[73] H. Zhang, I. Goodfellow, D. Metaxas, A. Odena, Self-attention generativeadversarial networks, Proceedings of the 36th International Conference onMachine Learning (2019) 7354{7363.
[74] Y. Zhou et al., LAFITE: Towards language-free training for text-to-imagegeneration, arXiv preprint arXiv:2111.13792 (2021).
[75] J-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image transla-tion using cycle-consistent adversarial networks, Proceedings of the IEEEInternational Conference on Computer Vision (2017) 2223{2232.