APPLICATIONS OF GENERATIVE ADVERSARIAL NETWORKS IN IMAGES STYLE TRANSFER: A SURVEY

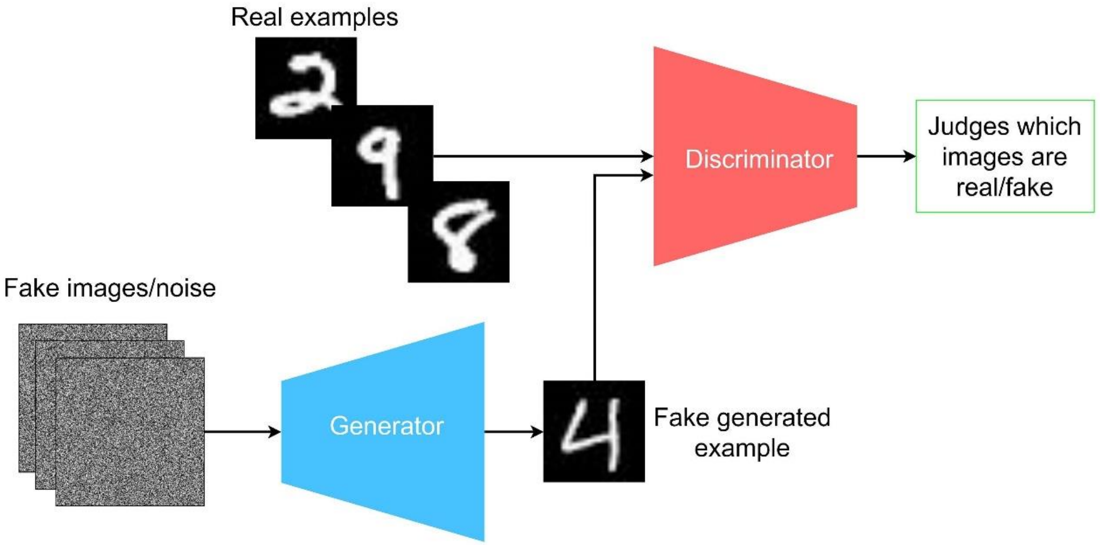
LIVIU-ȘTEFAN NEACȘU-MICLEA

Why?

- GANs can be powerful automated drawing tools
- Better adaptability than traditional software editors
- Save time for animators/illustrators
- Studies on style transfer can reveal important discoveries about features disentanglement and how to manipulate distributions
- Unpaired image-to-image style transfer
  - Useful when collecting paired data is too expensive
  - It relies on domain distribution instead of individual samples

Generative Adversial Networks

- Generator (G) and Discriminator (D)
- G tries to “fool” D into thinking counterfeit samples are real
- Really hard to train (How to train a GAN?)

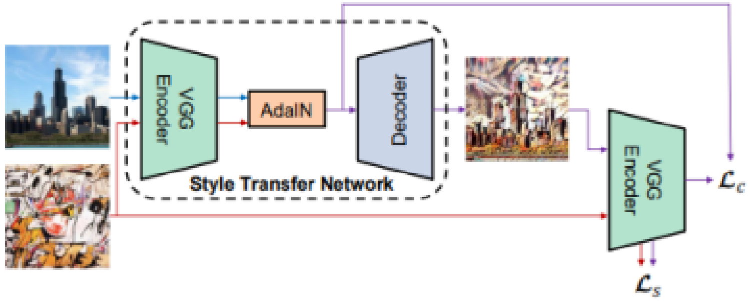


(Neural) Style Transfer

Input	Output
- Content Image (C) - Style Image (S)	Image taking content from (C) drawn in the style of (S)

- Gatys et al. proved that **Content and Style are separable** with CNN
- NST = Style Transfer with NN
- Adaptive Instance Normalization (AdaIN):
  - Transform a distribution to take the mean and variance of anoter distribution
  - Useful when we want to mix style with content

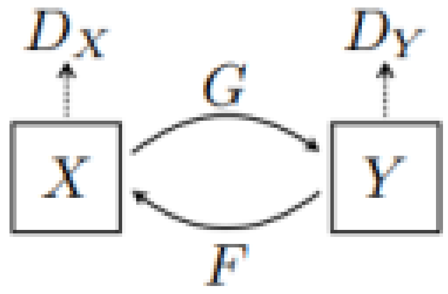
$$AdaIN(x,y) = \sigma(y)\left(\frac{x-\mu(x)}{\sigma(x)}\right) + \mu(y)$$



Architecture of AdaIN style transfer architecture

Neural Style Transfer with AdaIN

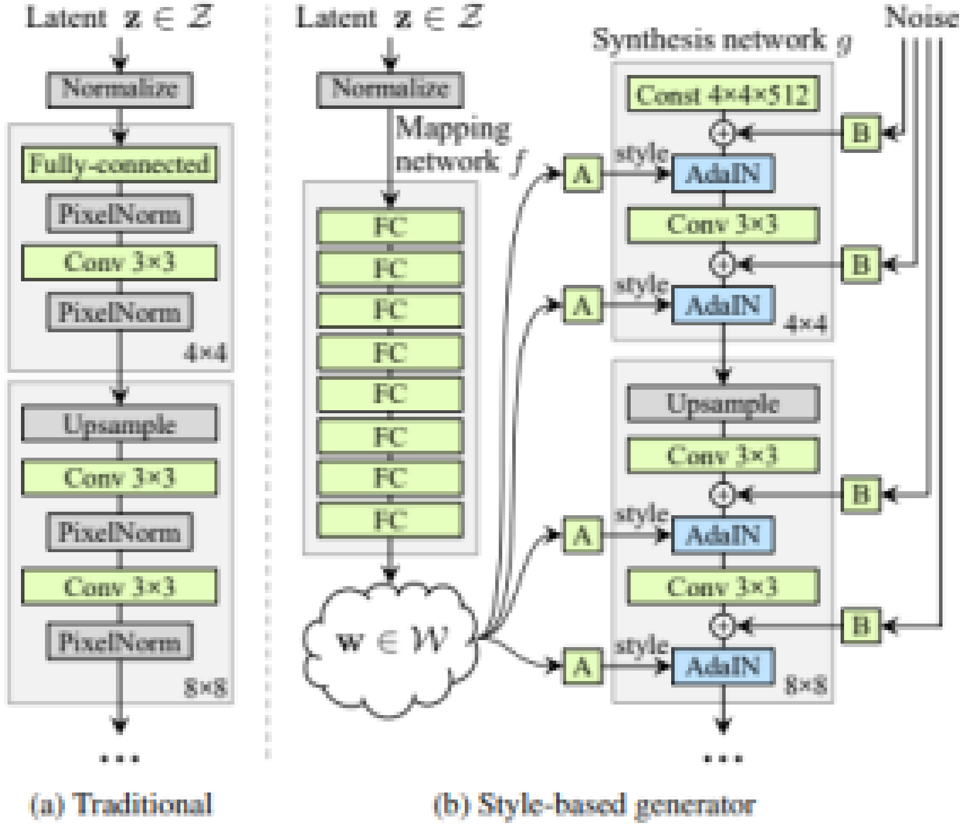
CycleGAN



CycleGAN architecture

Cycle consistency

StyleGAN

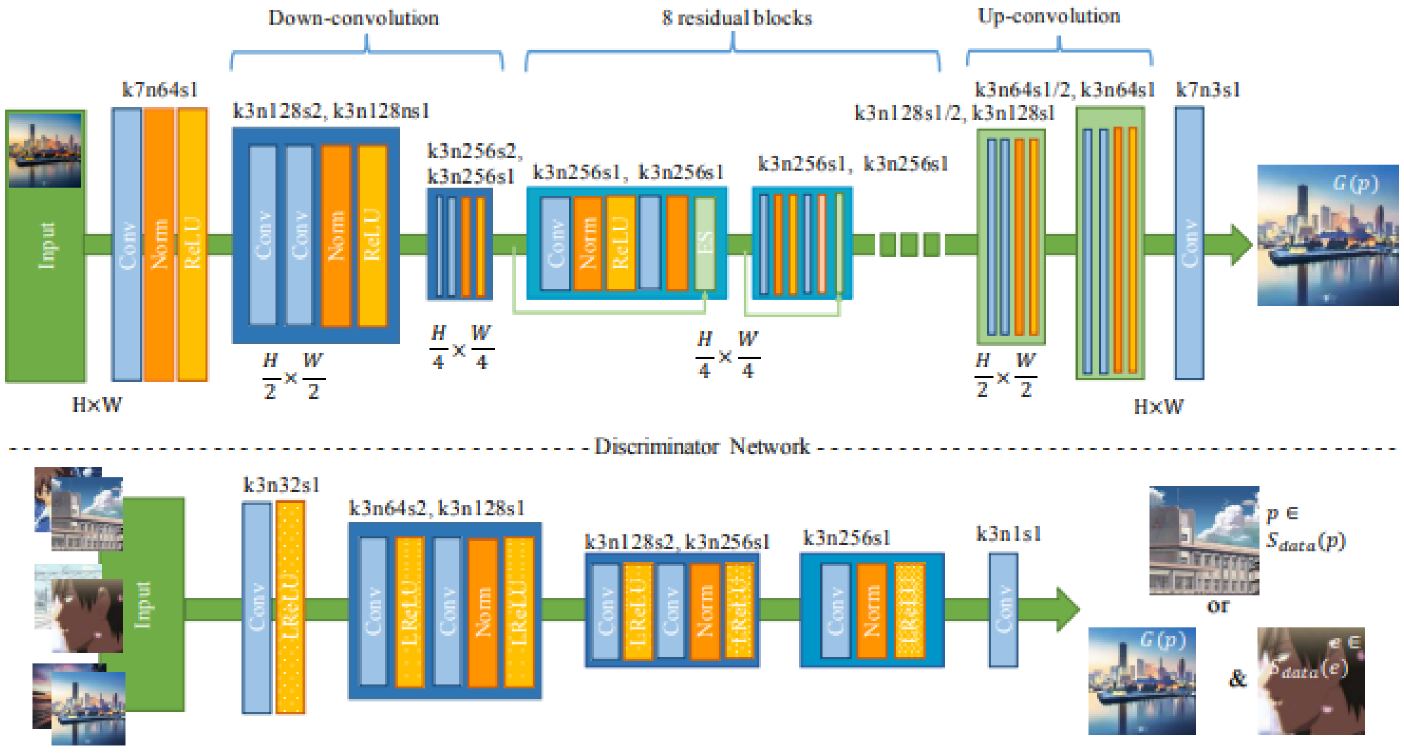


StyleGAN architecture

Stochastic variation

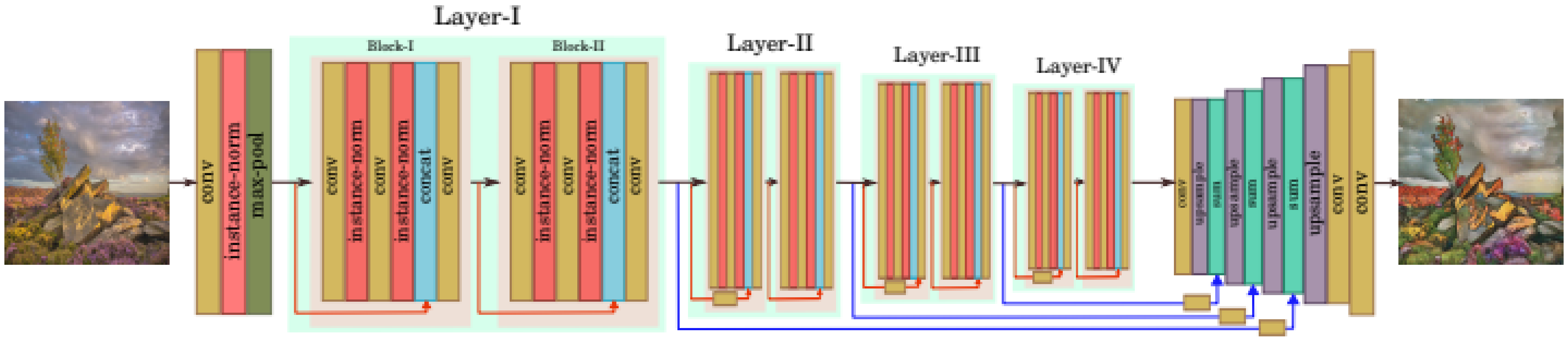
CartoonGAN

- Lightweight alternative to cycle-consistency
- Adapted for simplicity and edginess of cartoons
- Discriminator trained on real, fake and edge blurred images
- Pretrain generator on identity function as initialization step
  - Boosts training convergence
- Andersson's Miyazaki GAN
  - Direct implementation of CartoonGAN



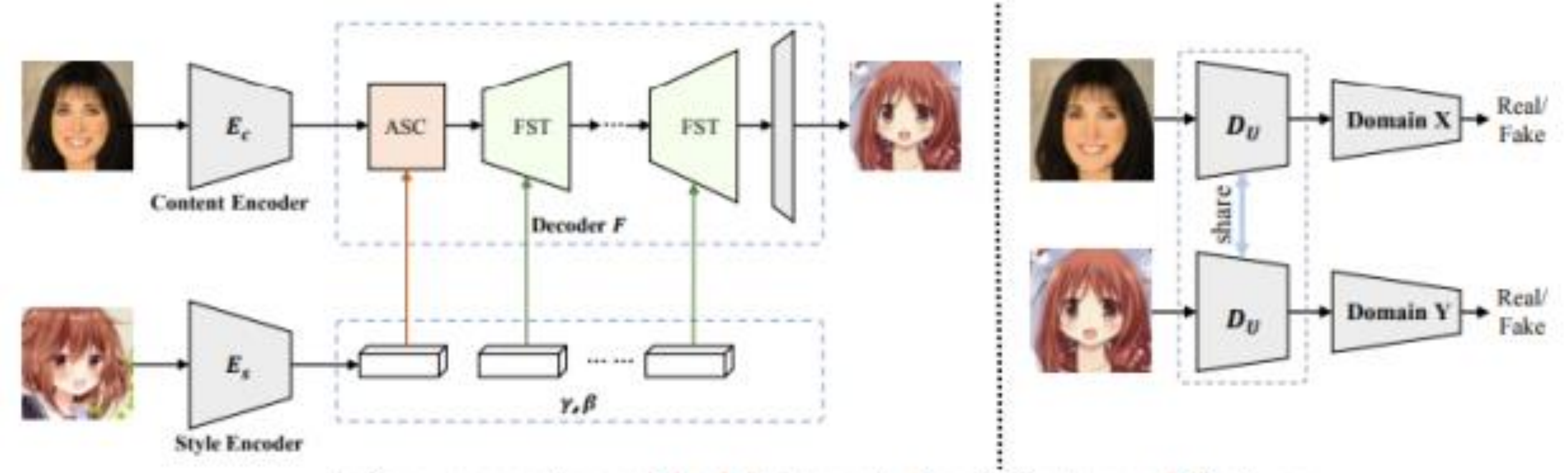
GANILLA

- Designed for book illustrations style transfer
- ResNet-18 based generator with skip-connections to upsampler
- Trained on works of 10 individual artists



AniGAN

- Previous models have problems with image geometry translations
- AniGAN specializes itself in face to anime style transfer tasks
- Complex generator (Adaptive Stack Convolutional (ASC), FineGrained Style Transfer (FST) blocks, AdaPoLIN)
- “Siamese”-like discriminator with shared weights to capture common structure of faces



Anigan generator and double branch discriminator architecture

GAN evaluation metrics

- Qualitative analysis
  - User study – Subjective evaluation, questionaries
- Quantitative analysis
  - Inception Score (IS)
    - passes generated images through a pretrained InceptionV3 and measure distribution's entropy, classification confidence and distribution
  - Fréchet inception distance
    - compares statistics of distributions of generated images with a real (ground truth) set
  - FCN-score
    - Used by CycleGAN in mask-to-picture task
  - Content & Style CNN metrics
    - CNN classifiers are pretrained to detect the presence of image content/style of an artist. The accuracy of the models across all generated samples is used as a metric (introduced by GANILLA paper)
  - Image-specific metrics: SSIM, Peak signal-to-noise ratio (PSNR)

Experimental setups in literature

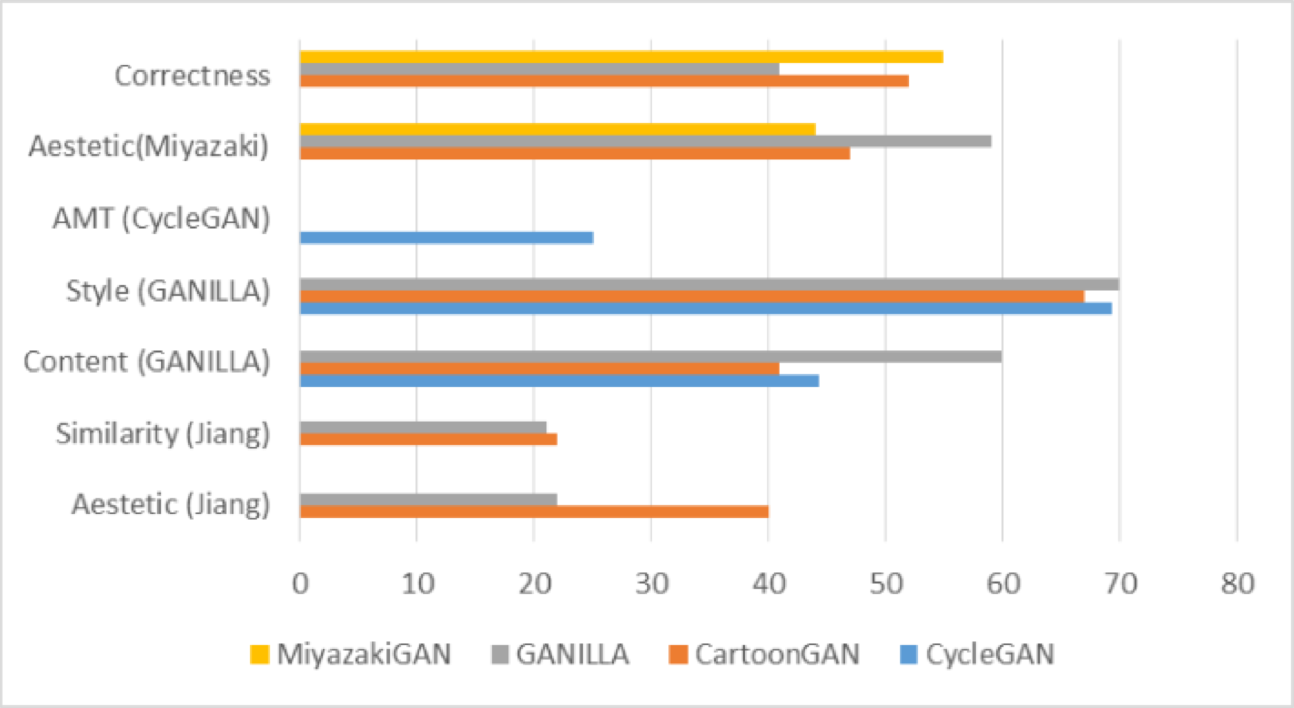
	CycleGAN [2]	StyleGAN [2]	CartoonGAN	GANILLA	MiyazakiGAN
Dataset(content)	Horse2Zebra, Summer2Winter	FFHQ*, LSUN	Flickr scraps*	CycleGAN's	Flickr30K
Dataset(style)			M. Shinkai & M. Hosoda art*	illustrators' work*	H. Miyazaki art*
Epochs	200	1000	200	200	60
Learning rate	$2 \cdot 10^{-4}$	$1 \cdot 10^{-3}$	$2 \cdot 10^{-4}$ **2	$2 \cdot 10^{-4}$	$1 \cdot 10^{-3}$ **1
Batch size	1	8	16 **2	1	11 **1
GPUs (NVIDIA)	2x Tesla V100	4x Tesla V100	Titan XP	Tesla V100	RTX 2080
No.pms.[7]	11.4mil	26.2mil[12]	11.1mil	7.2mil	N/A
Train.time[7]	1347s	1 week[12]	1400s	887s	144h[1]

Evaluation metrics in literature

	CycleGAN	StyleGAN	CartoonGAN	GANILLA	MiyazakiGAN
User Study	Illus.[7], M2A[15], O/CST[11]	-	Illus.[7], F2S/H*[4], MiyaST[1]	Illus.[7], O/CST[11], MiyaST[1]	MiyaST[1]
Style/Cont. CNN	Illus.[7]	-	Illus.[7]	Illus.[7]	-
FCN-score [9]	cityscapes[15]	-	-	-	-
FID	H2Z,S2W[2]	FFHQ[2],[12]	-	-	-
IS	H2Z,S2W[2]	FFHQ[2]	-	-	-
SSIM	O/CST[11]	-	O/CST[11]	O/CST[11]	-
PSNR	O/CST[11]	-	O/CST[11]	O/CST[11]	-

Metrics and datasets used to evaluate the models, found in literature
Datasets: M2A=map2aerial, H2Z=horse2zebra, S2W=summer2winter,
Illus.=GANILLA's illustrations dataset, F2S/H=Flickr to Shinkai/Hayao style,
O/CST = oil and cartoon style transfer datasets [11]
* = visual comparison, no numerical data; - = no data

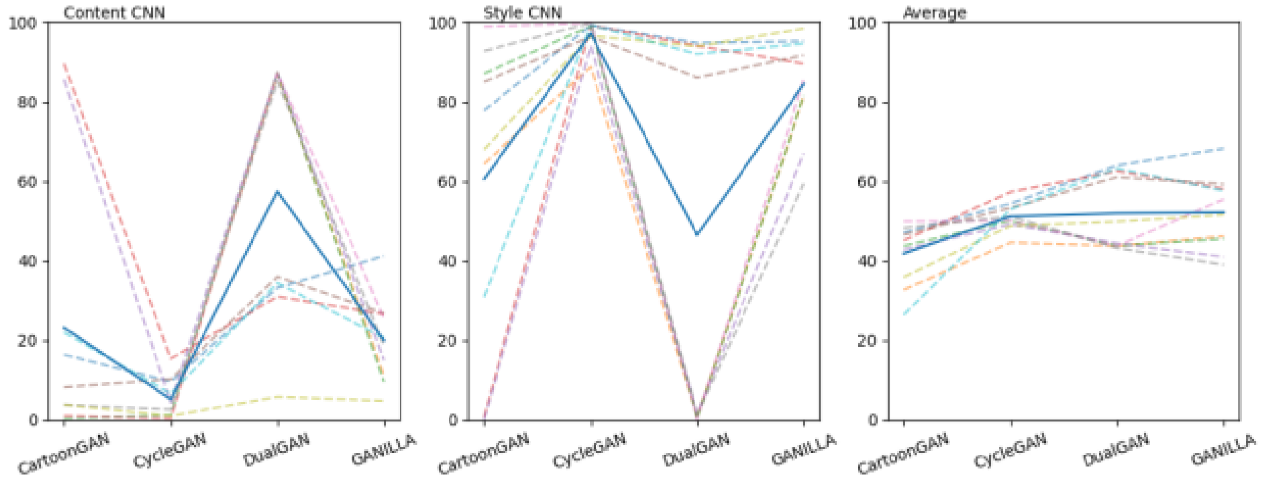
Evaluation statistics (Qualitative)



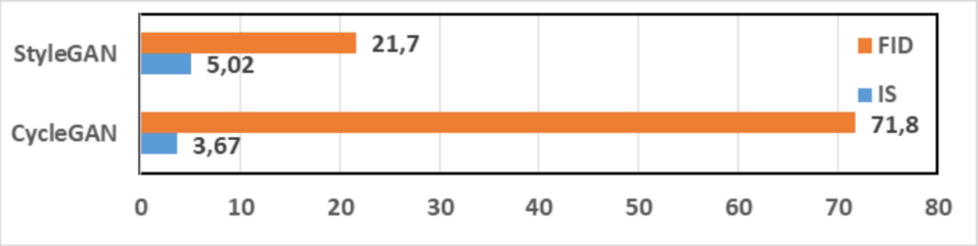
User studies conducted on the GAN architectures

Evaluation statistics (Qualitative)

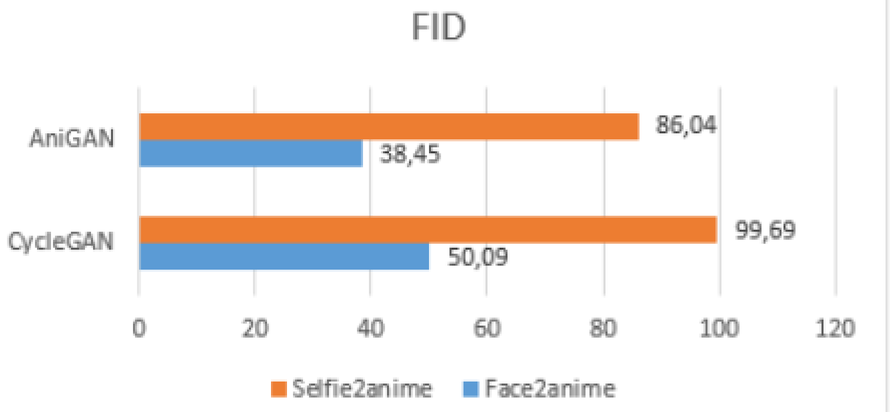
Evaluation statistics (Quantitative)



CNN metrics comparison between different GAN architectures
dashed lines = measure per illustrator style; continuous line = average measures per entire dataset
(GANILLA's illustrators dataset)



Quantitative comparison between StyleGAN and CycleGAN [2]



AniGAN quantitative metrics

Conclusions

- Impressive ideas, designs, performances
- Sometimes even with lower computational cost (CartoonGAN)
- There is no “good for all” method
  - Every problem is designed to solve its own style transfer proble (cartoonization, oil painting, illustration)
- ST encourages advancements in representation learning (StyleGAN, AniGAN)
- Improvements of the survey:
  - Try effectively run the discussed models ourselves
  - Unify the evaluations wherever possible (same metrics, same datasets)
  - What about other GANs? VAE? Diffusion? …

Thank you for your attention!

References

Bibliography:

- Gatys, L., Ecker, A., Bethge, M., 2016. A neural algorithm of artistic style. Journal of Vision 16, 326–326. doi:10.1167/16.12.326
- Karras, T., Laine, S., Aila, T., 2019. A style-based generator architecture for generative adversarial networks, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4401–4410.
- Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE international conference on computer vision, pp. 2223–2232.
- Hicsonmez, S., Samet, N., Akbas, E., Duygulu, P., 2020. Ganilla: Generative adversarial networks for image to illustration translation. Image and Vision Computing 95, 103886.
- Chen, Y., Lai, Y.K., Liu, Y.J., 2018. Cartoongan: Generative adversarial networks for photo cartoonization, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9465–9474. doi:10.1109/CVPR.2018.00986.

Images:

- https://ro.pinterest.com/pin/593067844662779311/
- https://developer.ibm.com/articles/generative-adversarial-networks-explained/
- https://www.kaggle.com/code/stefanneacsu/catart-style-transfer-adain