Photo Realistic WCT

- Use exact WCT architecture as before
  - …but use max un-pooling in upsample layers, instead of transpose convolutions (*meh*)
  - …and a smoothing constraint applied as an optimization on the result
  - Notation is borrowed from graph manifold rankings:

$$\arg\min_{R} \frac{1}{2} \sum_{i,j \in C}^{N,M} e^{-\frac{\|I_i^c - I_j^c\|^2}{\sigma_{i,j}^2}} \left\| \frac{R_i}{\sqrt{D_{ii}}} - \frac{R_j}{\sqrt{D_{jj}}} \right\|^2 + \left(\frac{1}{\alpha} - 1 \right) \sum_{i}^{N} \sum_{j}^{M} \| R_{i,j} - Y_{i,j} \|^2$$

Y. Li, M.-Y. Liu, X. Li, M.-H. Yang, J. Kautz, A Closed-form Solution to Photorealistic Image Stylization, 2018

Smoothing

$\mathbf{I}^c$ is the content image, $\mathbf{Y}$ is the stylized image in graph structure

$\mathbf{R}$ is the desired result in graph structure

weighted sum of adjacent pixels i, j in $\mathbf{R}$ | mixing hyperparameter | L2 norm of graphs

$$\arg\min_R \frac{1}{2} \boxed{\sum_{i,j\in 1\Delta}^{N,M} \boxed{e^{-\frac{\|I_i^c - I_j^c\|^2}{\sigma_{i,j}^2}}} \left\| \frac{R_i}{\sqrt{D_{ii}}} - \frac{R_j}{\sqrt{D_{jj}}} \right\|^2} + \boxed{\left(\frac{1}{\alpha} - 1\right)} \boxed{\sum_i^N \sum_j^M \left\| R_{i,j} - Y_{i,j} \right\|^2}$$

$\mathbf{W}_{i,j}$

affinity of content image as graph edges
normalized by std of neighboring pixels (1Δ)
known as "Matting Affinity"

$$D_{ii} = \sum_{\forall j} e^{-\frac{\|I_i^c - I_j^c\|^2}{\sigma_{i,j}^2}} = \sum_{\forall j} W_{i,j}$$

D is a diagonal matrix (degree matrix), summed from W

$$\hat{\mathbf{R}} = (1-\alpha)(\mathbf{I} - \alpha \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}})^{-1} \mathbf{Y}$$

closed form solution for smoothed result (I is identity matrix)

Y. Li, M.-Y. Liu, X. Li, M.-H. Yang, J. Kautz, A Closed-form Solution to Photorealistic Image Stylization, 2018

What is W? Connectivity of pixels as graph

$$D_{ii} = \sum_{\forall j} e^{-\frac{\|I_i^c - I_j^c\|^2}{\sigma_{i,j}^2}} = \sum_{\forall j} W_{i,j}$$

$\mathbf{I}^c$

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

$$W_{6,7} = e^{-\frac{(I_6 - I_7)^2}{\sigma_{1\rightarrow 11}^2}}$$

1	2	3
5	6	7
9	10	11

sigma neighborhood

$W_{i,j}$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	Δ			Δ	Δ	Δ									
2	Δ	0	Δ			Δ										
3		Δ	0	Δ		Δ	Δ	Δ								
4			Δ	0			Δ	Δ								
5	Δ				0	Δ			Δ	Δ						
6	Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ					
7		Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ				
8			Δ	Δ			Δ	0			Δ	Δ				
9					Δ	Δ			0	Δ			Δ	Δ		
10					Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ	
11						Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ
12							Δ	Δ			Δ	0			Δ	Δ
13									Δ	Δ			0	Δ		
14									Δ	Δ	Δ		Δ	0	Δ	
15										Δ	Δ	Δ		Δ	0	Δ
16											Δ	Δ			Δ	0

sum
over
each
row

D_{ii}

What is W, Y, and R?

$$\underset{N^2xN^2}{\hat{\mathbf{R}}} = (1-\alpha)\boxed{(\underset{\substack{N^2xN^2\\ \text{identity}}}{\mathbf{I}} - \alpha \underset{N^2xN^2}{\mathbf{D}^{-\frac{1}{2}}} \underset{N^2xN^2}{\mathbf{W}} \underset{N^2xN^2}{\mathbf{D}^{-\frac{1}{2}}}}) ^{-1} \underset{N^2xN^2}{\mathbf{Y}}$$

Laplacian of graph

$W_{i,j}$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	Δ			Δ	Δ	Δ									
2	Δ	0	Δ			Δ										
3		Δ	0	Δ		Δ	Δ	Δ								
4			Δ	0			Δ	Δ								
5	Δ				0	Δ			Δ	Δ						
6	Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ					
7		Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ				
8			Δ	Δ			Δ	0			Δ	Δ				
9					Δ	Δ			0	Δ			Δ	Δ		
10					Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ	
11						Δ	Δ	Δ		Δ	0	Δ		Δ	Δ	Δ
12							Δ	Δ			Δ	0			Δ	Δ
13									Δ	Δ			0	Δ		
14									Δ	Δ	Δ		Δ	0	Δ	
15										Δ	Δ	Δ		Δ	0	Δ
16											Δ	Δ			Δ	0

sum over each row

D_{ii}

- **D** is diagonal and easily invertible
- **W** is sparse and efficiently inverted after multiplications
- **Y** is the stylized image pixels on a diagonal matrix
- **R** can be converted to an image by returning the diagonal

How to make this graph?

sklearn.feature_extraction.image.grid_to_graph

```
sklearn.feature_extraction.image.grid_to_graph(n_x, n_y, n_z=1, mask=None, return_as=<class 'scipy.sparse.coo.coo_matrix'>, dtype=<class 'int'>)
```

[source]

Graph of the pixel-to-pixel connections

Edges exist if 2 voxels are connected.

Parameters:

n_x : ***int***
Dimension in x axis

n_y : ***int***
Dimension in y axis

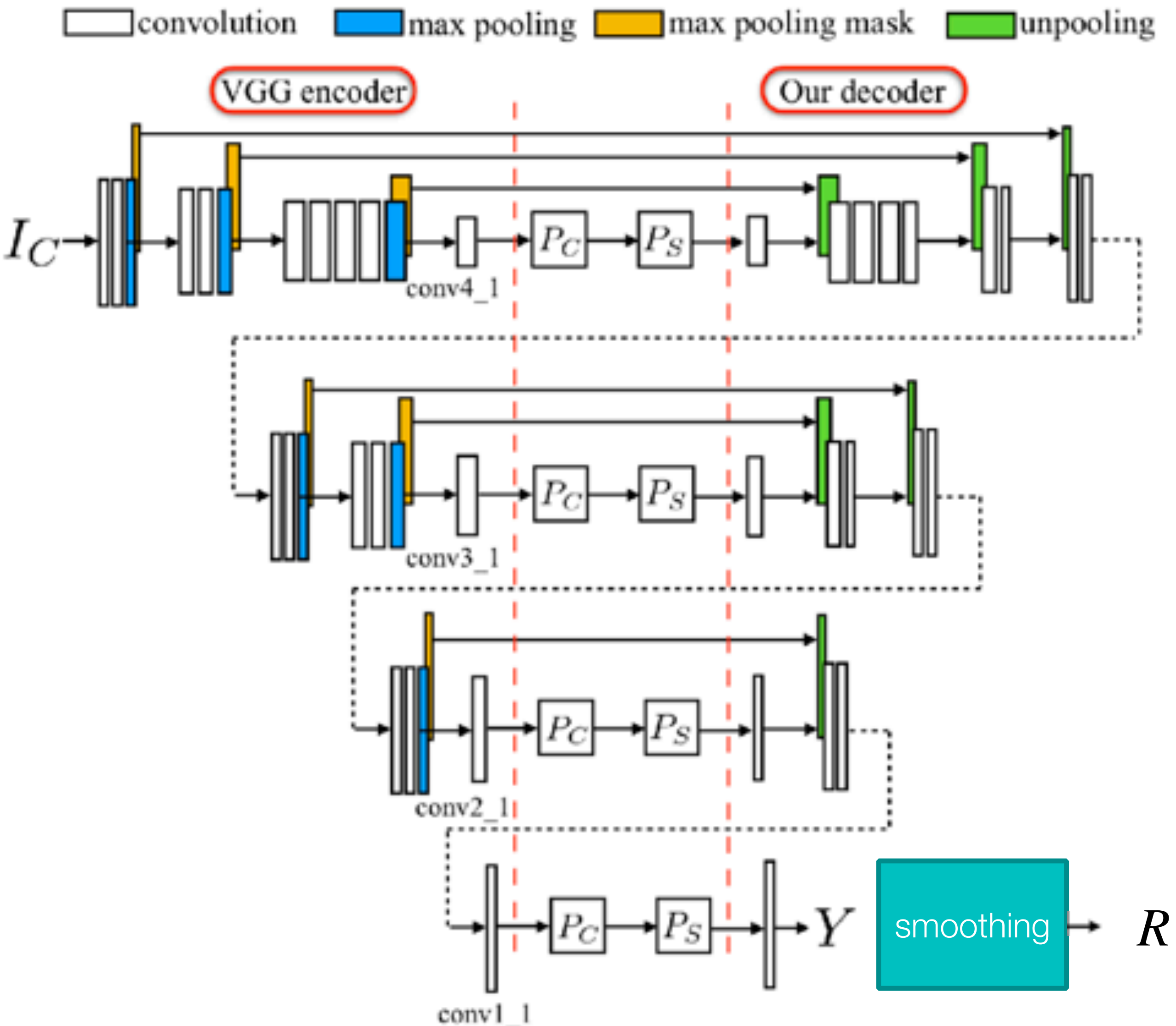
n_z : ***int, optional, default 1***
Dimension in z axis

mask : ***ndarray of booleans, optional***
An optional mask of the image, to consider only part of the pixels.

return_as : ***np.ndarray or a sparse matrix class, optional***
The class to use to build the returned adjacency matrix.

dtype : ***dtype, optional, default int***
The data of the returned sparse matrix. By default it is int

Similar Architecture as Before



Results

(a) Style (b) Content

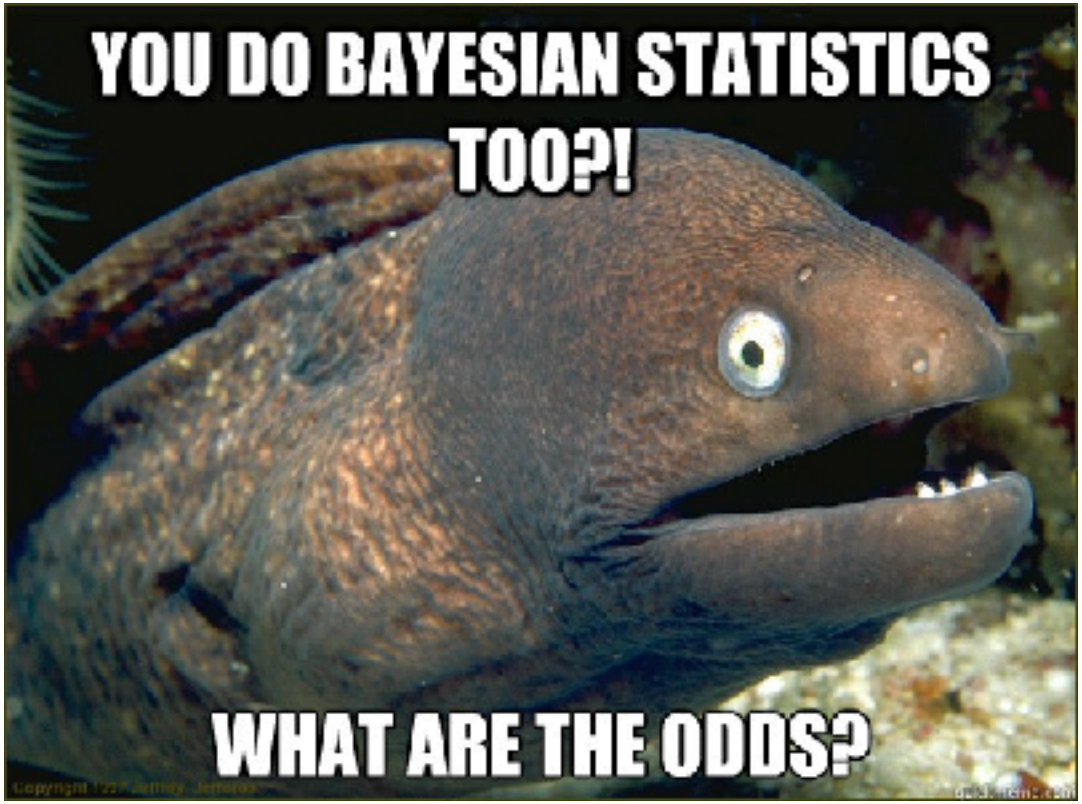
(c) WCT [10] (d) PhotoWCT

(e) WCT + smoothing (f) PhotoWCT + smoothing

Apply Masking to Different Segments of Image

Paper	Loss	Description
Gatys et al. [4]	*Gram Loss*	The first proposed style loss based on Gram-based style representations.
Johnson et al. [43]	*Perceptual Loss*	Widely adopted content loss based on perceptual similarity.
Berger and Memisevic [29]	*Transformed Gram Loss*	Computing *Gram Loss* over horizontally and vertically translated features. More effective at modelling style with symmetric properties, compared with *Gram Loss*.
Li et al. [51]	*Mean-substraction Gram Loss*	Subtracting the mean of feature representations before computing *Gram Loss*. Eliminating large discrepancy in scale. Effective at multi-style transfer with one single network.
Zhang and Dana [52]	*Multi-scale Gram Loss*	Computing *Gram Loss* over multi-scale features. Eliminating a few artefacts.
Li et al. [38]	*MMD Loss with Different Kernels*	*Gram Loss* is equivalent to *MMD Loss with Second Order Polynomial Kernel*. *MMD Loss with Linear Kernel* is capable of comparable quality with *Gram Loss*, but with lower computational complexity.
Li et al. [38]	*BN Loss*	Achieving comparable quality with *Gram Loss*, but conceptually clearer in theory.
Risser et al. [40]	*Histogram Loss*	Matching the entire histogram of feature representations. Eliminating instability artefacts, compared with single *Gram Loss*.
Li et al. [41]	*Laplacian Loss*	Eliminating distorted structures and irregular artefacts.
Li and Wand [42]	*MRF Loss*	More effective when the content and style are similar in shape and perspective, compared with *Gram Loss*.
Champandard [65]	*Semantic Loss*	Incorporating a segmentation mask over *MRF Loss*. Enabling a more accurate match.
Gu et al. [54]	*Reshuffle Loss*	Connecting both global and local style losses. Capable of preserving global appearance while avoiding distortions in local style patterns.
Li and Wand [48]	*Adversarial Loss*	Computed based on PatchGAN. Utilising contextual correspondence between patches. More effective at preserving coherent textures in complex images.
Jing et al. [61]	*Stroke Loss*	Achieving continuous stroke size control while preserving stroke consistency.
Wang et al. [62]	*Hierarchical Loss*	Enabling a coarse-to-fine stylisation procedure. Capable of producing large but also subtle strokes for high-resolution content images.
Liu et al. [63]	*Depth Loss*	Preserving depth maps of content images. Effective at retaining spatial layout and structure of content images, compared with single *Gram Loss*.
Ruder et al. [72]	*Temporal Consistency Loss*	Designed for video style transfer. Penalising the deviations along point trajectories based on optical flow. Capable of maintaining temporal consistency among stylised frames.
Chen et al. [70]	*Disparity Loss*	Designed for stereoscopic style transfer. Penalising bidirectional disparity. Capable of consistent strokes for different views.

Evaluation

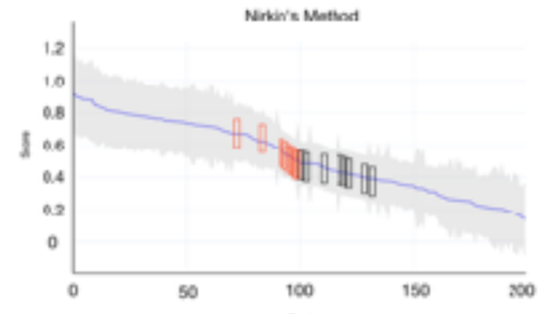


Evaluation

- Qualitative
  - Show a few cherry picked results in the paper
    - …preferred by researchers who care little about the merits of evaluation but need a publication
  - User testing:
    - **Single choice**: Pick preferred style from a list, which method has more votes
    - **Two-up random testing**: Force users to choose preferred style in two styled images, rank final images
    - **Infinite Rating Scale**: Place Images along a continuous rating scale, allowing infinite precision (my Masters Thesis), Allows for similarity measure

Fast, Can be Remote
Requires Many Observations for Convergence

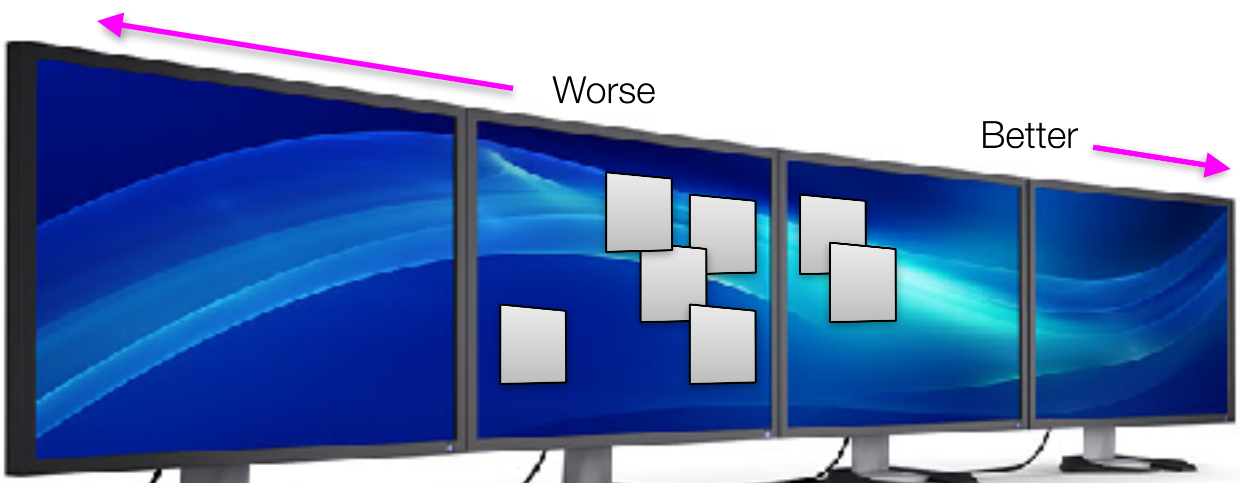
Can achieve approximate rankings



Swapped Face Detection using Deep Learning and Subjective Assessment

Xinyi Ding[1*], Zohreh Raziei[2], Eric C. Larson[1], Eli V. Olinick[3], Paul Krueger[3] and Michael Hahsler

Image Preference Ranking



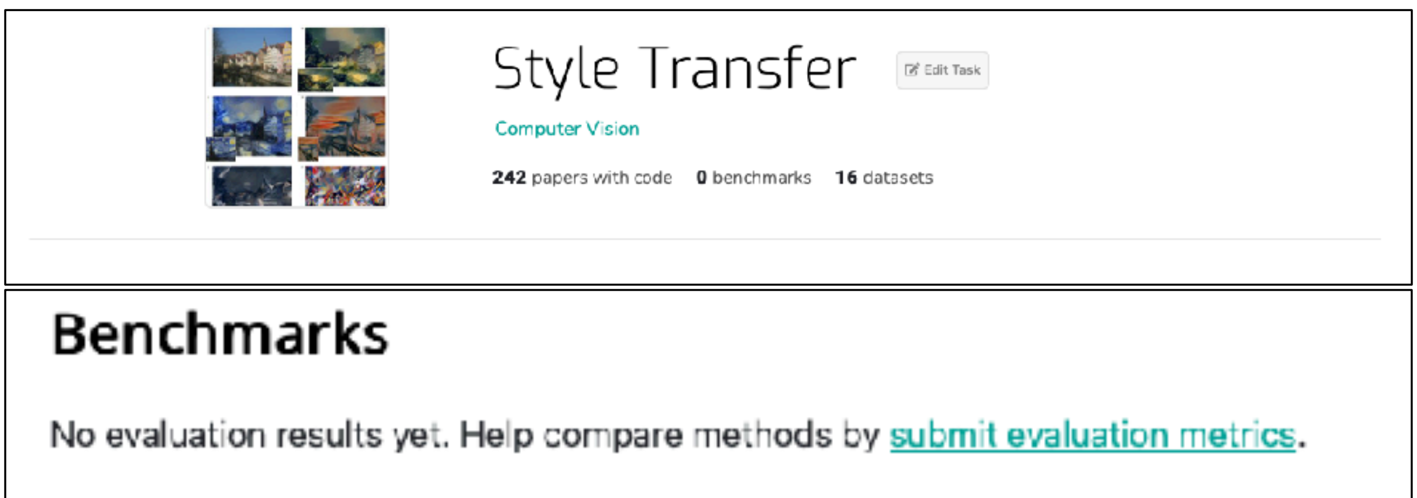
Evaluation from EC Larson Master's Thesis

The CSIQ images were subjectively rated based on a linear displacement of the images across four calibrated LCD monitors placed side-by-side with equal viewing distance to the observer. The database contains 5000 subjective ratings from 35 different observers. Ratings were corrected for personal bias using "agreement" scrolls of images.

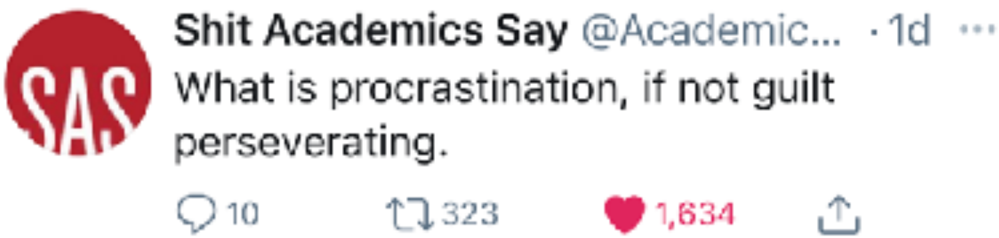
Evaluation

- Quantitative ¯_(ツ)_/¯

- There is no concrete, reliable quantitative measure
- That would be a human in the loop
- Any paper claiming their loss function is more minimized than others does not have respect for the perception of a human being's ability to discern art



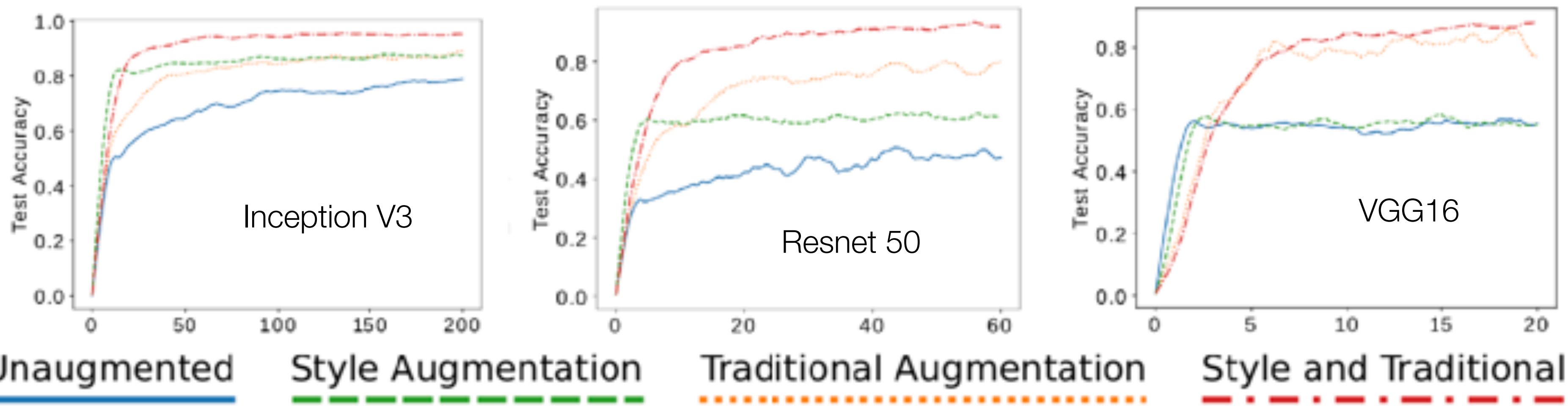
Style Transfer Applications and Other Domains



Data Augmentation

- Current image augmentation can make robust to spatial location, rotation, skew, *etc.*
- Perhaps style augmentation can achieve color, texture invariance?

Philip T. Jackson, Amir Atapour-Abarghouei, Stephen Bonner, Toby Breckon, Boguslaw Obara, 2018

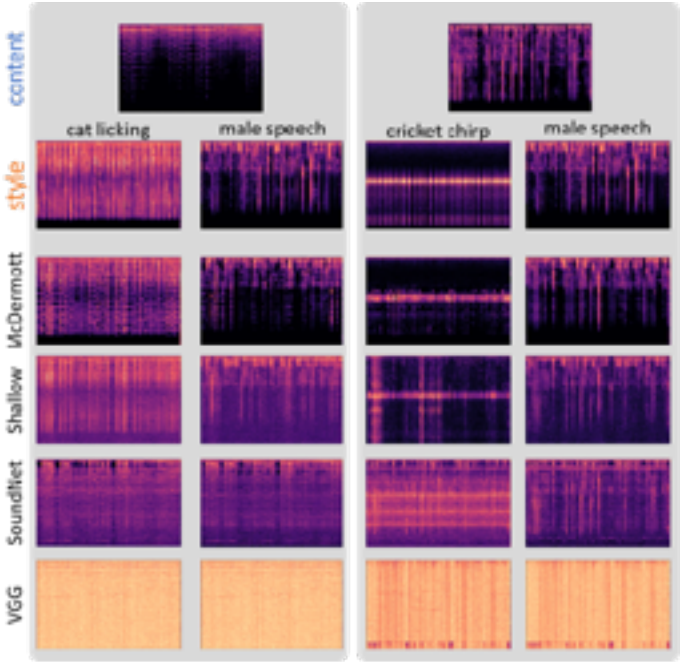


Current Applications

- Social Media: Image Filtering and Communication
- Architecture and Interior Design
- Digital Art and Photography (Adobe)
- Gaming/Movie Industry (NVIDIA)

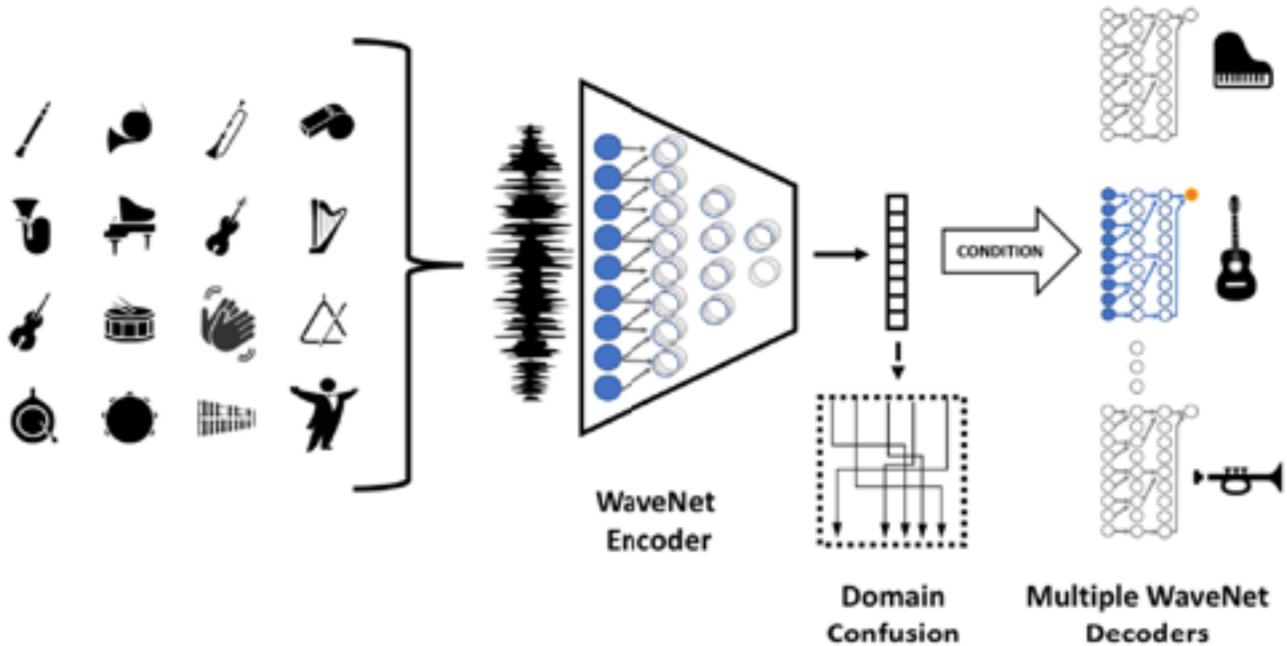
Audio Style Transfer

- Most works carry over on the audio spectrogram
- These works are lackluster, not really coherent
- Many open research questions
- One downside: most convolutions are handled as real numbers which makes little sense when applied the the STFT…



Audio Style Transfer with WaveNet

- WaveNet is an autoencoder for speech and music, capable of capturing many aspects of music from time domain samples
- FAIR Paper: Train single encoder, multiple decoders



Noam Mor, Lior Wolf, Adam Polyak, Yaniv Taigman, A Universal Music Translation Network , May 2018

State of the Art in Audio Transfer

- FAIR results are compelling…

Lecture Notes for

Neural Networks and Machine Learning

Style Transfer: One Shot, Other Domains

Next Time:
Transfer Learning

Reading: None