Style Transfer

*For queries contact : Kranthi Kumar R (ee18d004@smail.iitm.ac.in)

Style Transfer

- Given a piece of art and a photo, recreate the photo in the style of the art.

Content Image Style Image Stylized Result

The gradient of the score

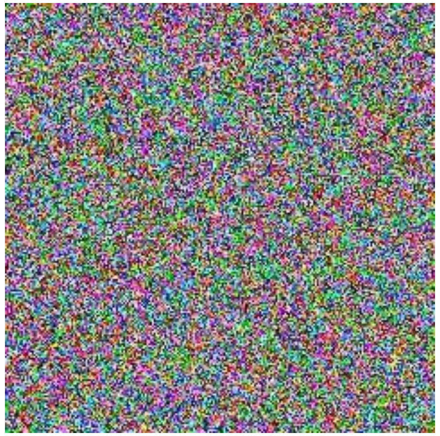
Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps.K. Simonyan, A. Vedaldi, A. Zisserman. ICLR Workshop 2014

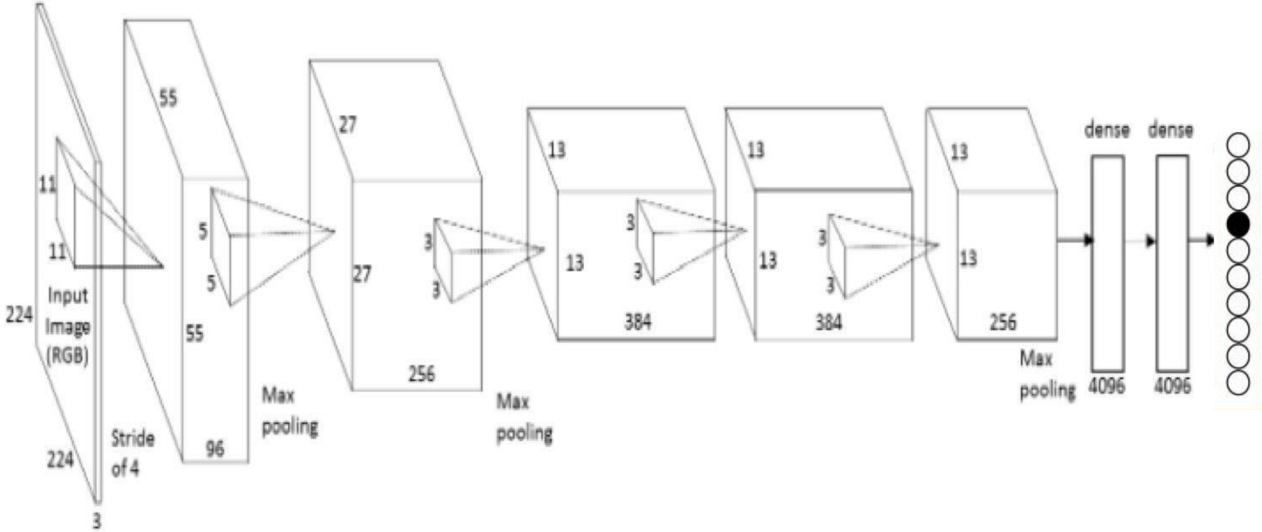
What input to a neuron maximizes a class score?

Neuron of choice *i* & An image of random noise *x*.

Repeat:

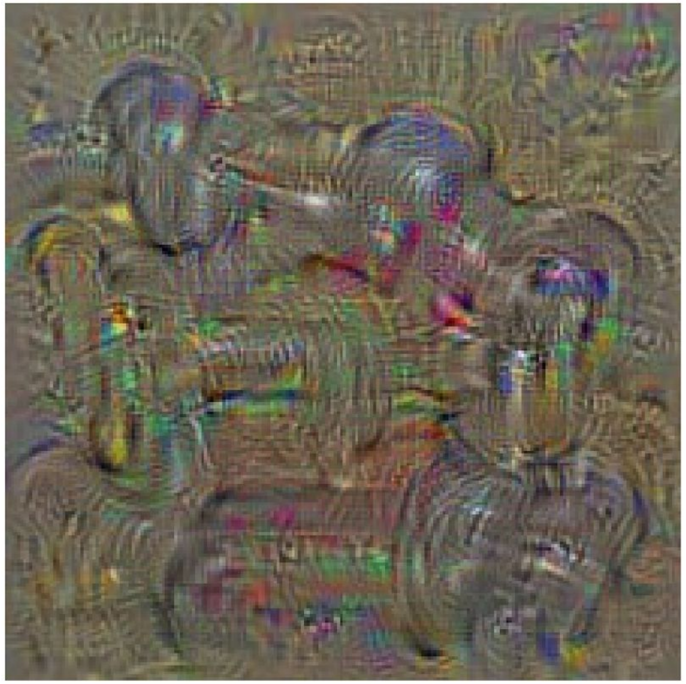
1. Forward propagate: compute activation $a_i(x)$
2. Back propagate: compute gradient at neuron $\partial a_i(x) / \partial x$
3. Add small amount of gradient back to noisy image.



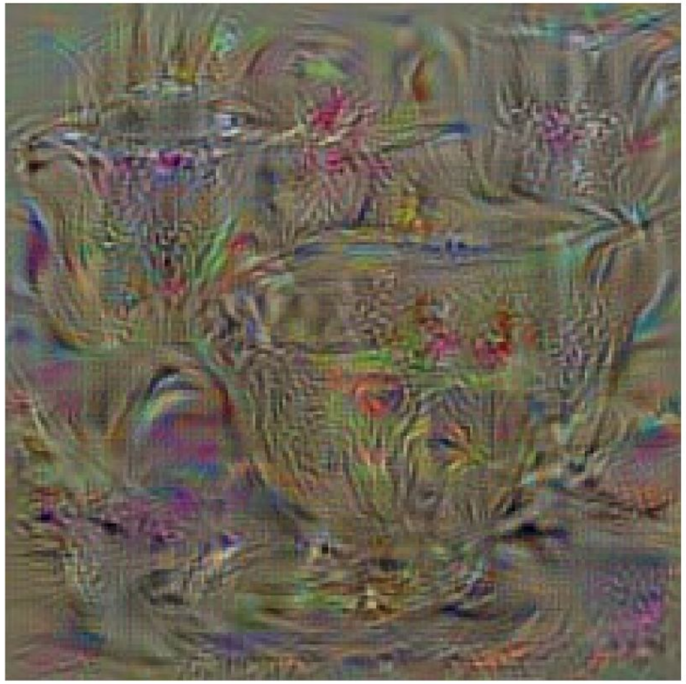


The image for a class

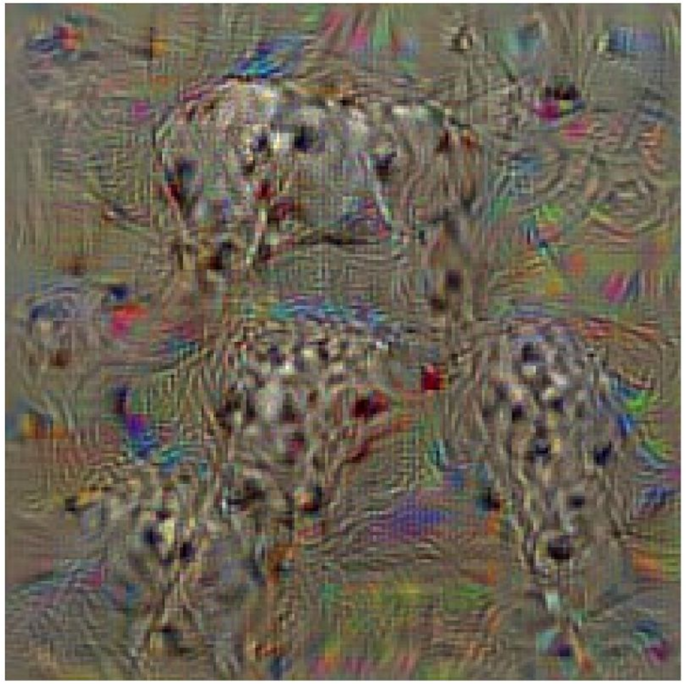
$$\arg\max_{I} S_c(I) - \lambda\|I\|_2^2,$$



dumbbell



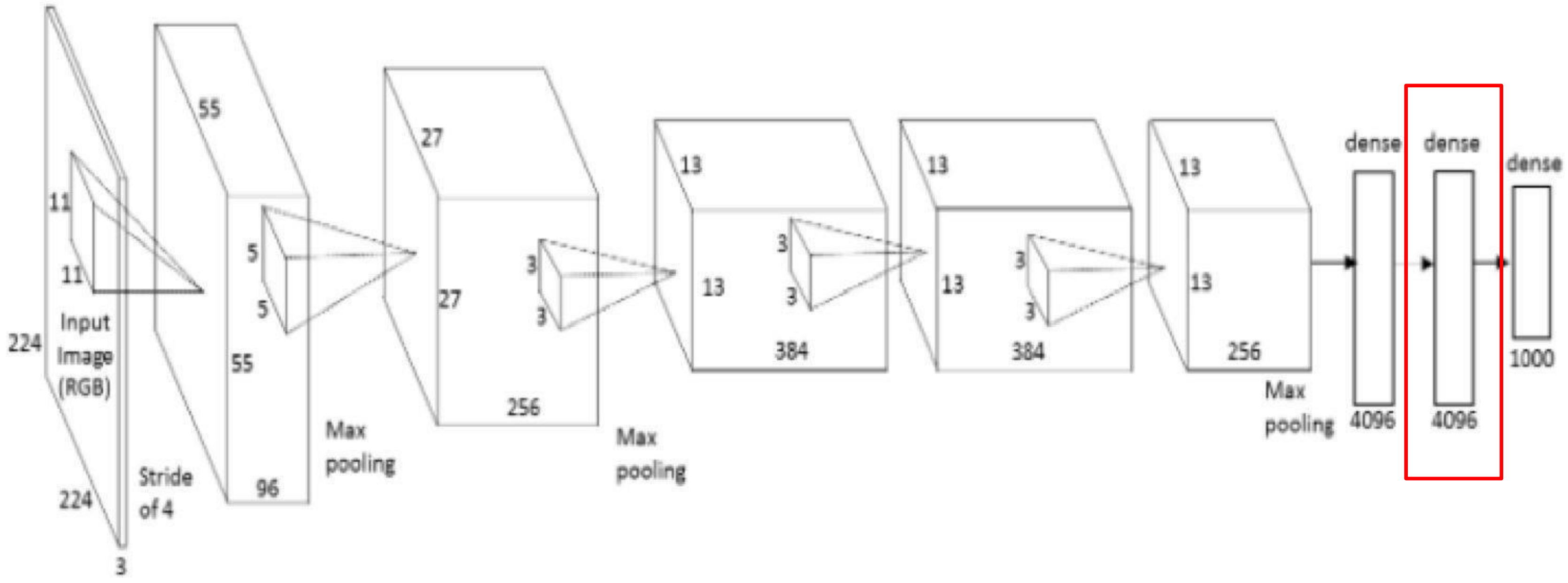
cup



dalmatian

Inverting convolutional networks

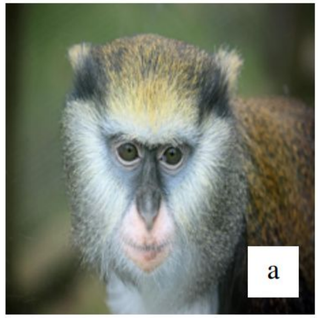
- Question: Given a CNN feature vector, is it possible to reconstruct the original image



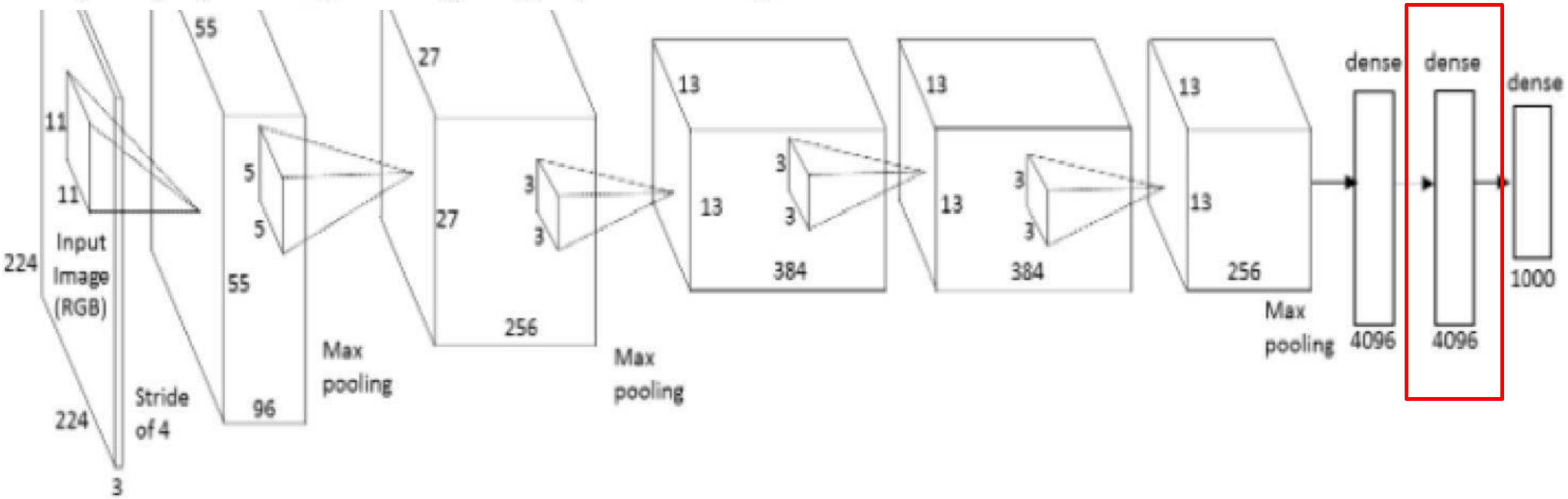
Inverting convolutional networks

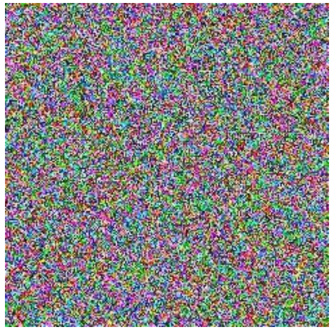
- Question: Given a CNN feature vector, is it possible to reconstruct the original image

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbb{R}^{H \times W \times C}}{\operatorname{argmin}} \ \ell(\Phi(\mathbf{x}), \Phi_0) + \lambda \mathcal{R}(\mathbf{x})$$

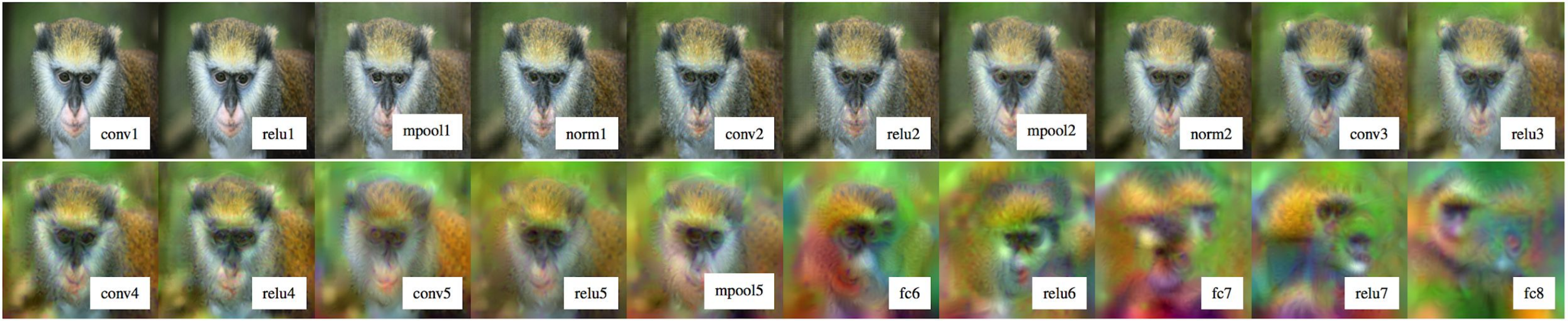


$$\ell(\Phi(\mathbf{x}), \Phi_0) = \|\Phi(\mathbf{x}) - \Phi_0\|^2$$





Inverting convolutional networks



Mahendran, Aravindh, and Andrea Vedaldi. "Understanding deep image representations by inverting them." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

Inverting convolutional networks: Texture

- Texture Matching – Statistical Property
  - Determine relevant texture properties (histograms, filter responses), synthesize image with matching properties
- Texture Representation:
  - Covariance matrix of texture properties
    - Spatially invariant
    - Average second-order statistics
  - *Gram matrices* ~ correlations between the different feature responses.

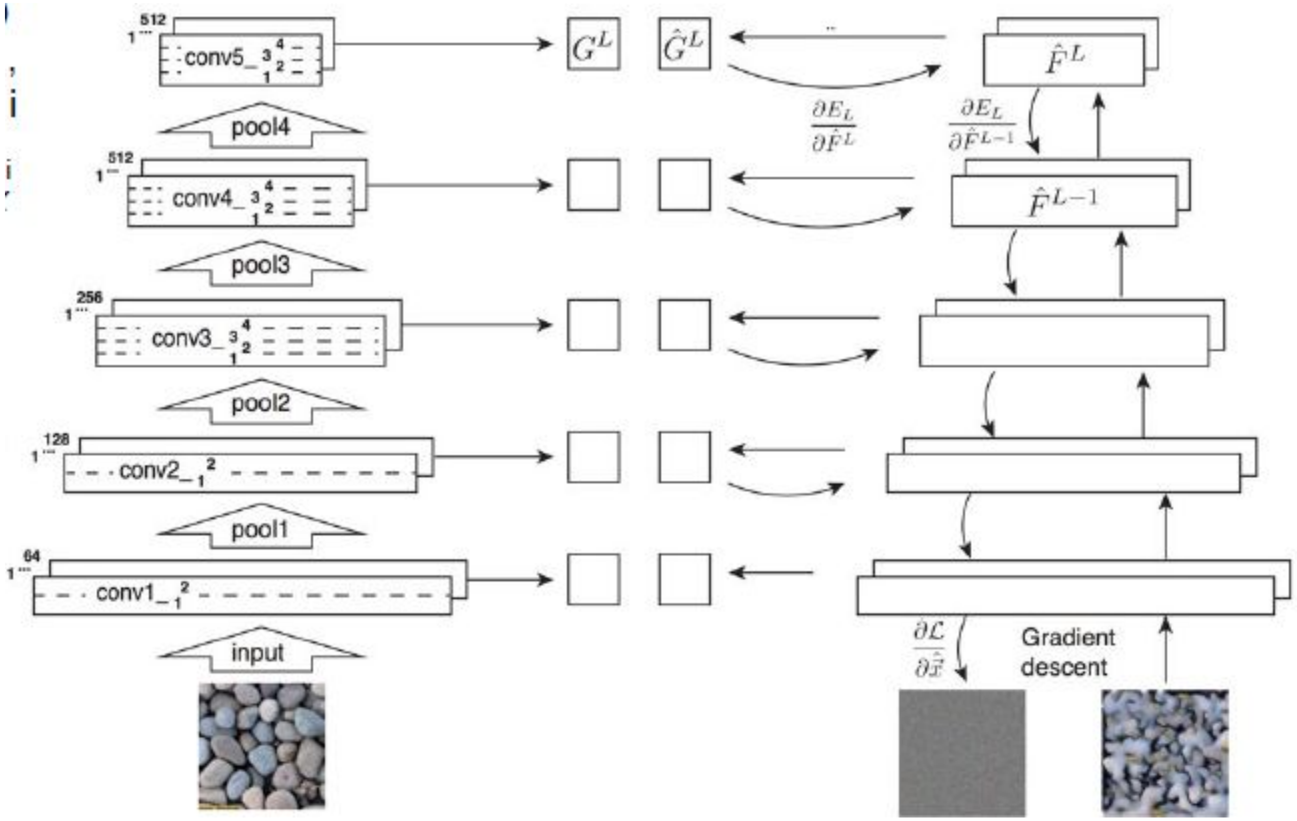
$$G = V^{\mathrm{T}} V$$

If dim(V) = [224x224x64]
then dim(G) = [64x64]

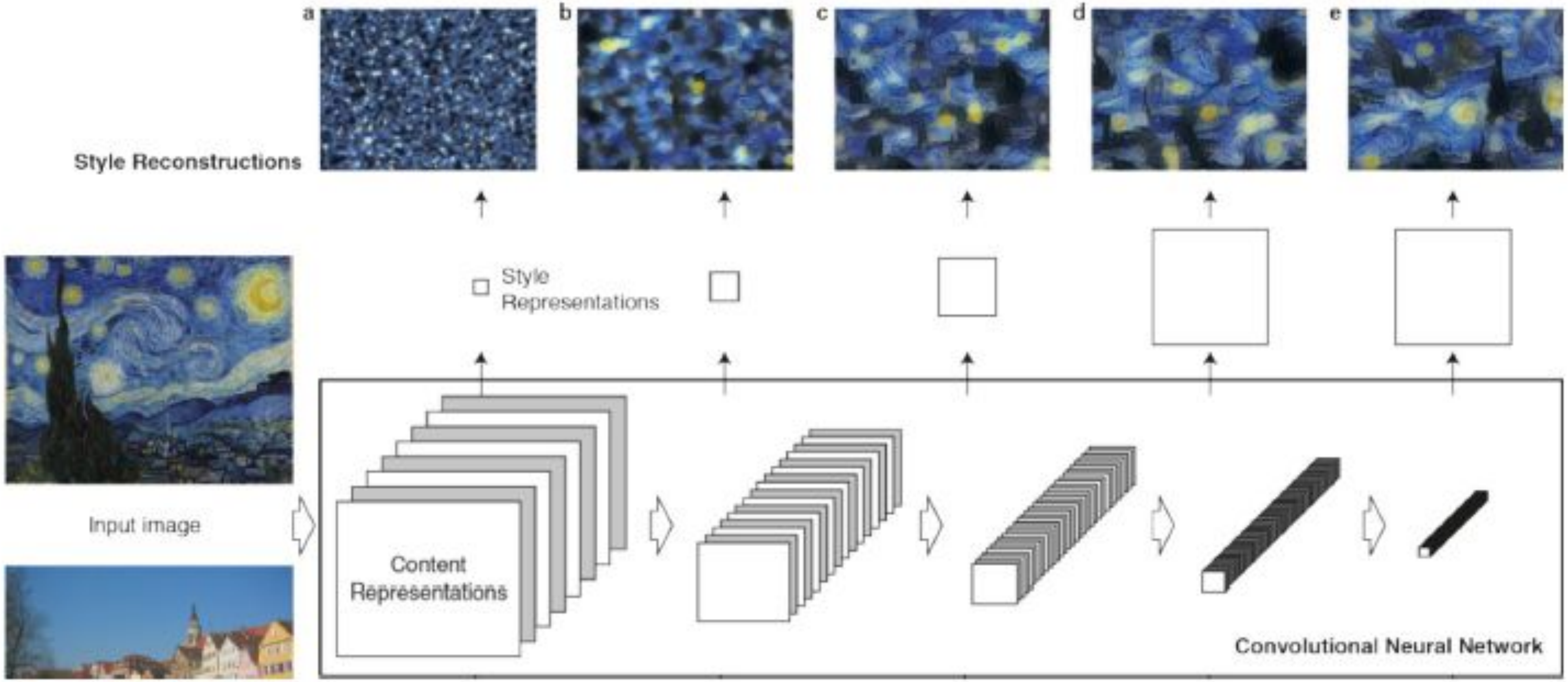
Gatys, Leon, Alexander S. Ecker, and Matthias Bethge. "Texture synthesis using convolutional neural networks." Advances in Neural Information Processing Systems. 2015

Inverting Convolutional networks: Texture

$$\mathbf{x}^* = \underset{\mathbf{x}\in\mathbb{R}^{H\times W\times C}}{\operatorname{argmin}} \ell(G^L, \hat{G}^L) + \lambda\mathcal{R}(\mathbf{x})$$



Inverting Convolutional networks: Texture



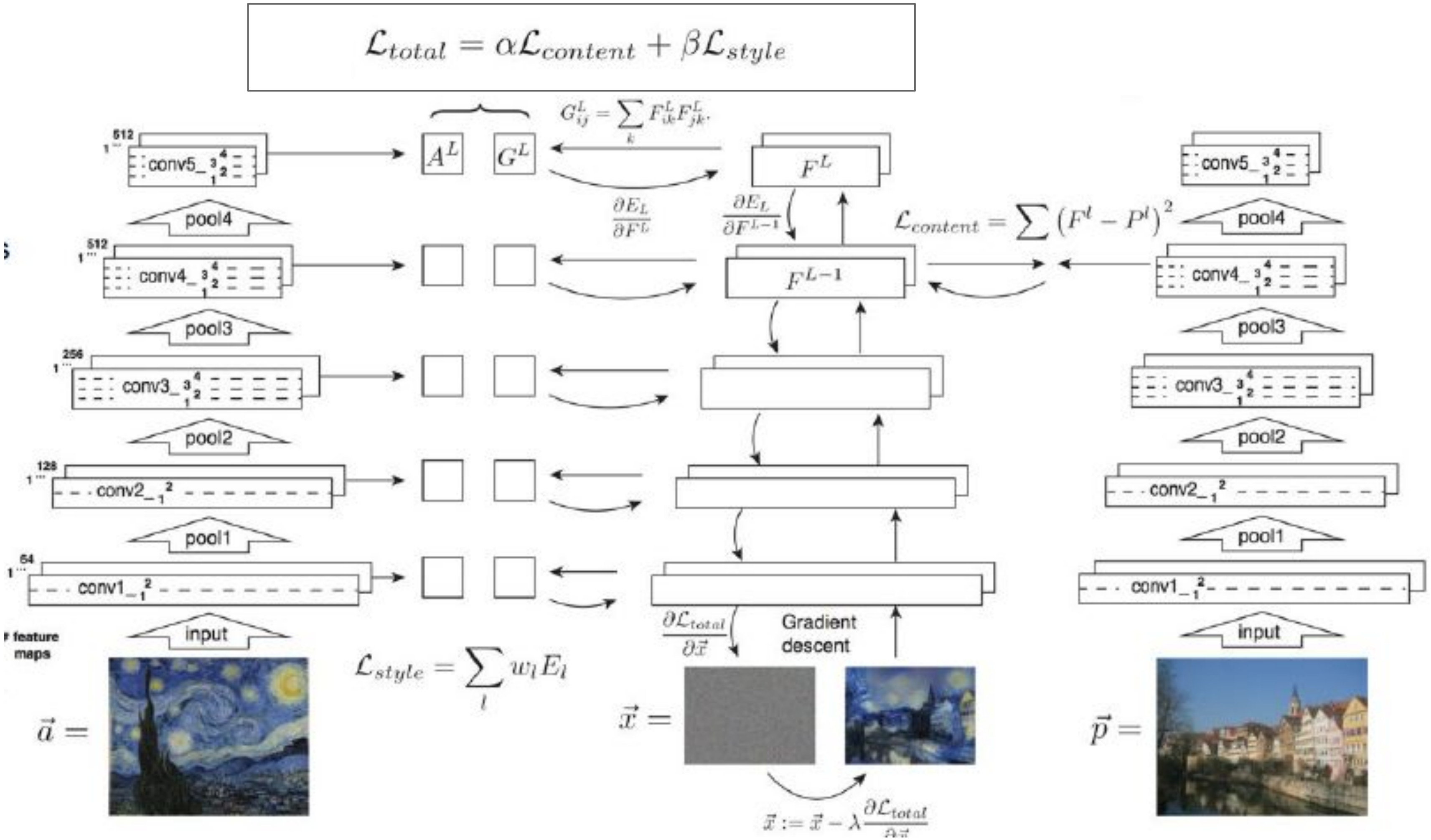
Side-effect - style transfer

- *Content representation:* feature map at each layer
- *Style representation:* Covariance matrix at each layer
    - Spatially invariant
    - Average second-order statistics

- Idea: Optimize x to match content of one image and style of another

Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." *arXiv preprint arXiv:1508.06576* (2015).

Neural Style transfer



Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "Image style transfer using convolutional neural networks." *CVPR* 2016

Neural Style Transfer

Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." *arXiv preprint arXiv:1508.06576* (2015).

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{content} + \beta \mathcal{L}_{style}$$

CNN-MRF

$$\tilde{\mathcal{L}}_s(\mathbf{x}, \mathbf{a}) = \sum_{i=1}^{m_l} \|\Psi_i(\Phi^l(\mathbf{x})) - \Psi_{NN(i)}(\Phi^l(\mathbf{a}))\|^2$$

- Different Style Representation
  - Correlations and Gram matrices ⟶ local patch based

Content Image Gatys et al Ours

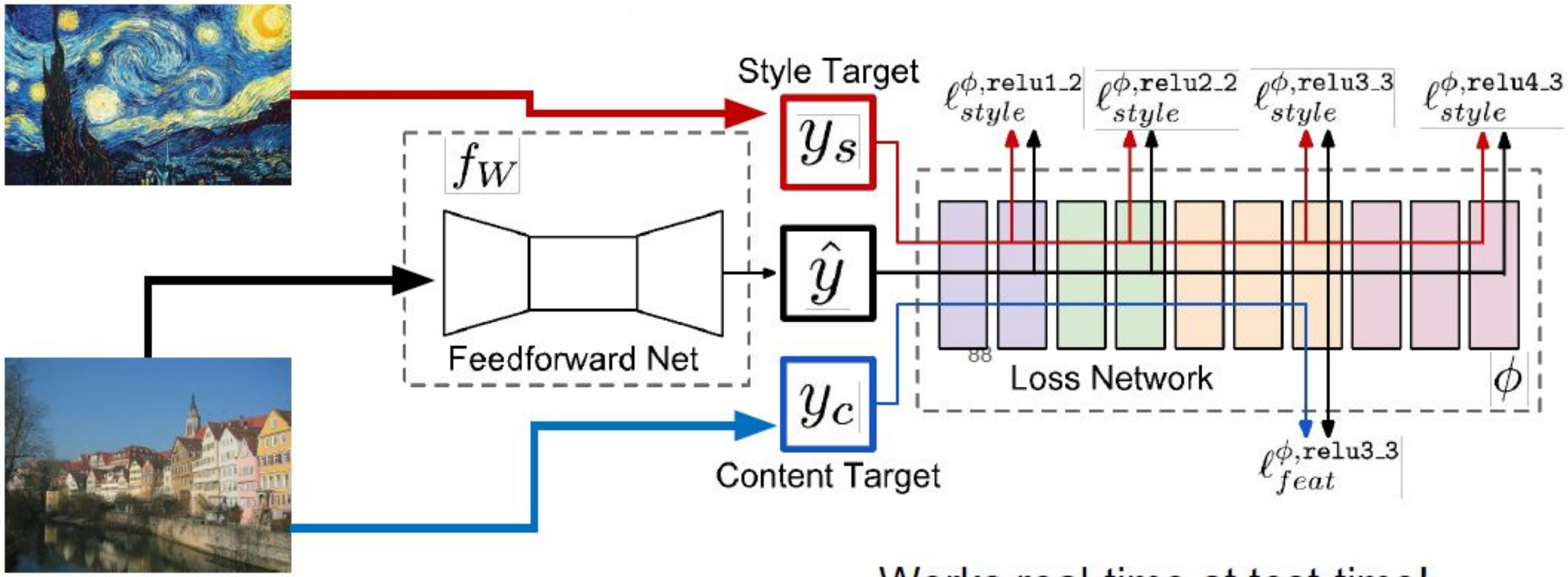
Content Image Gatys et al Ours

Input style

Input content Gatys et al Ours

(Artistic) Fast Style Transfer (1)

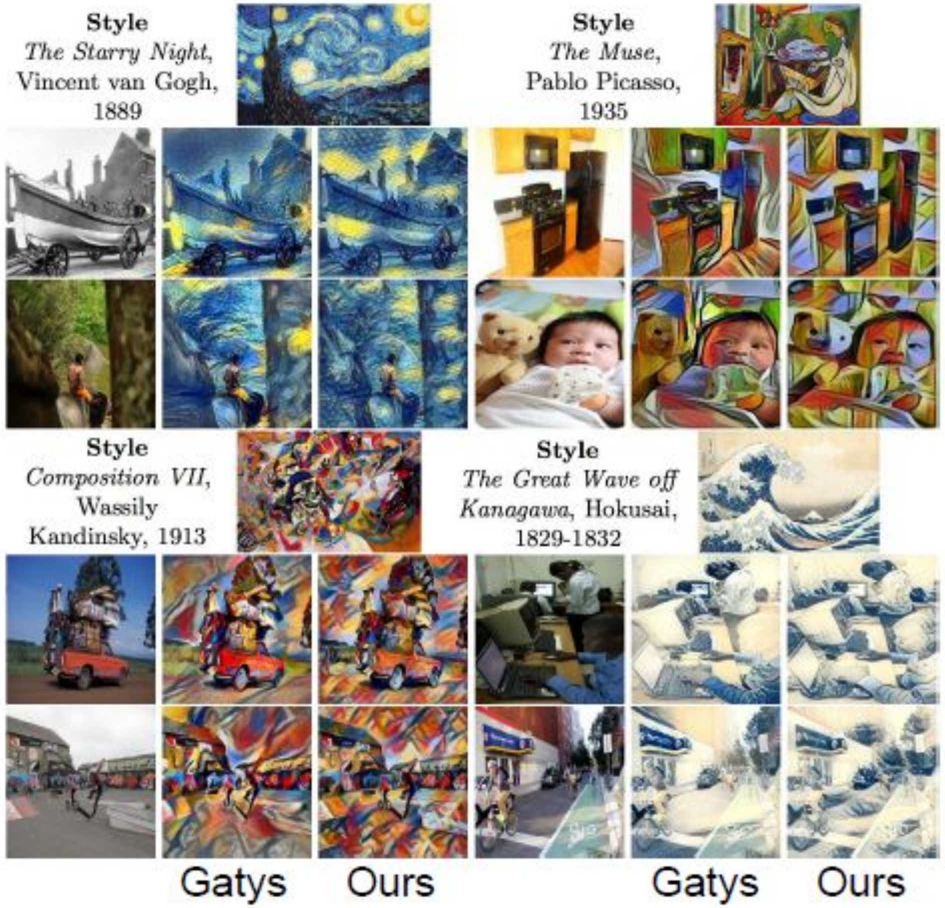
- Learning to transfer style
  1. Train a feedforward network for *each style*
  2. Use pre-trained CNN to compute same losses as before
  3. After training, stylize images using a single forward pass



Works real-time at test-time!

Perceptual Losses for Real-Time Style Transfer and Super-Resolution
Justin Johnson, Alexandre Alahi, Li Fei-Fei
ECCV 2016

(Artistic) Fast Style Transfer (1)

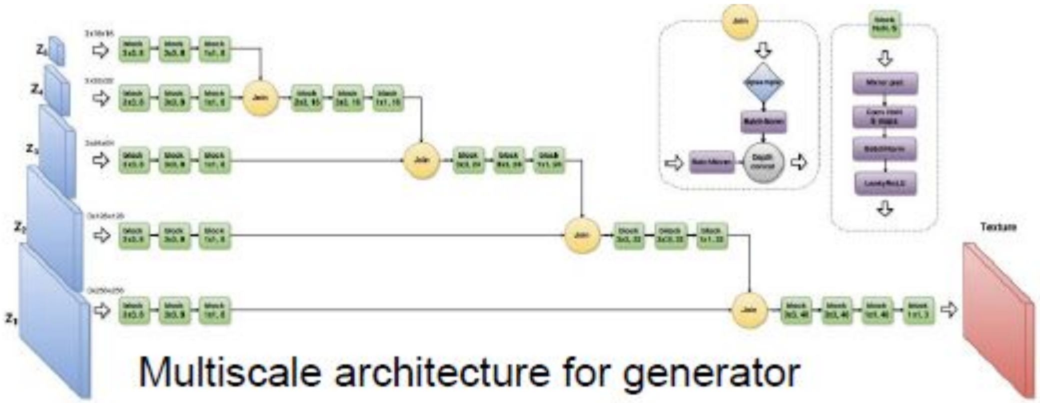


Perceptual Losses for Real-Time Style Transfer and Super-Resolution
Justin Johnson, Alexandre Alahi, Li Fei-Fei
ECCV 2016

(Artistic) Fast Style Transfer (2)

Ulyanov et al, "**Texture Networks: Feed-forward Synthesis of Textures and Stylized Images**", ICML 2016

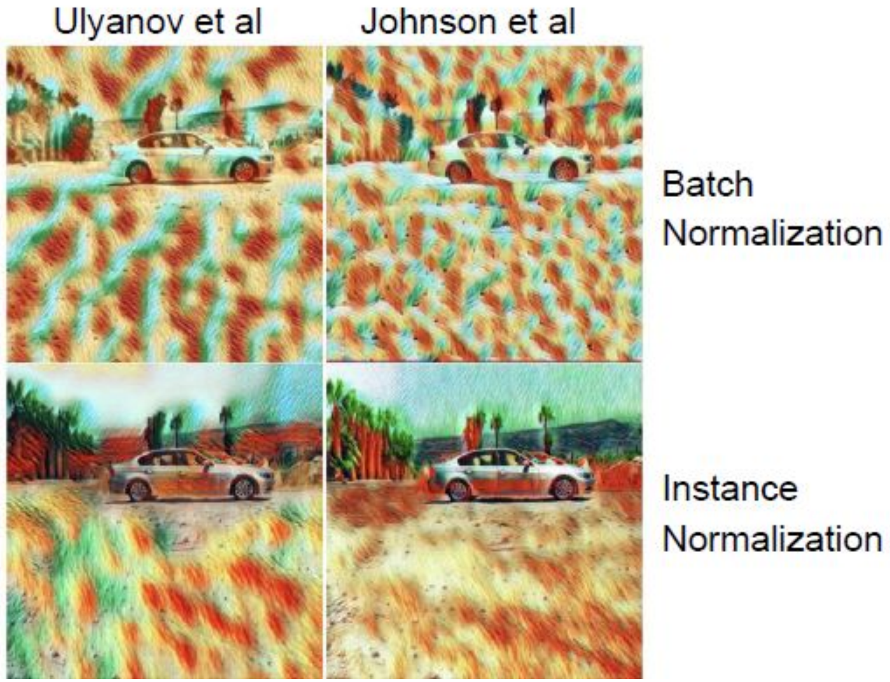
- Multi-scale architecture for generator
- Almost similar but ours 500 times faster.



Multiscale architecture for generator

Ulyanov et al, "**Instance Normalization: The Missing Ingredient for Fast Stylization**", ICML 2016

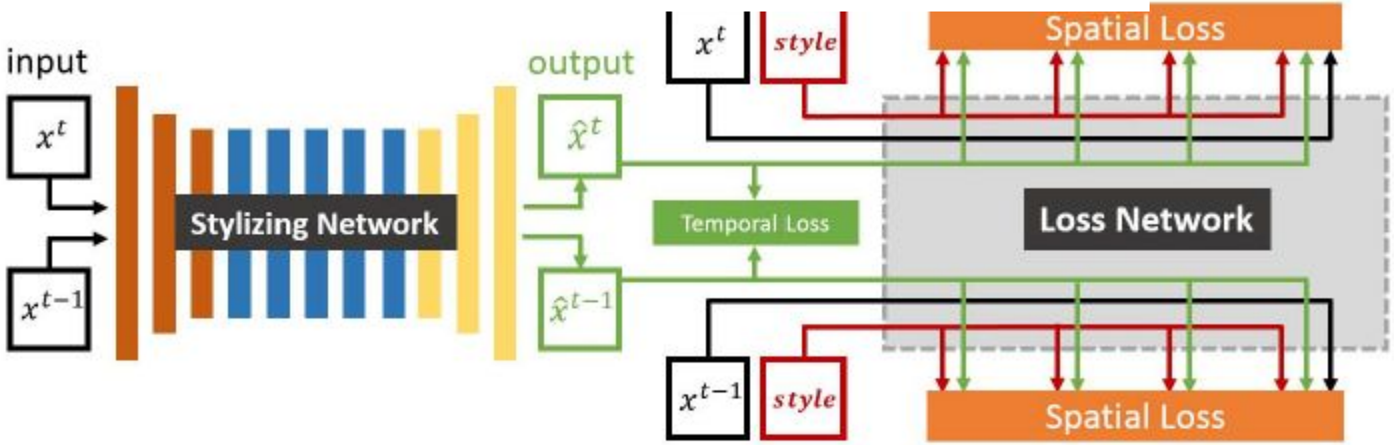
- Minor Tweak: Use instance Normalization instead of Batch Normalization



(Artistic) Style Transfer for Videos (1)

$$\mathcal{L}_{hybrid} = \underbrace{\sum_{i \in \{t,t-1\}} \mathcal{L}_{spatial}(\mathbf{x}^i, \hat{\mathbf{x}}^i, \mathbf{s})} + \underbrace{\lambda \mathcal{L}_{temporal}(\hat{\mathbf{x}}^t, \hat{\mathbf{x}}^{t-1})}_{\text{temporal loss}}$$

$$\mathcal{L}_{spatial}(\mathbf{x}^t, \hat{\mathbf{x}}^t, \mathbf{s}) = \underbrace{\alpha \sum_l \mathcal{L}^l_{content}(\mathbf{x}^t, \hat{\mathbf{x}}^t)}_{\text{content loss}} + \underbrace{\beta \sum_l \mathcal{L}^l_{style}(\mathbf{s}, \hat{\mathbf{x}}^t)}_{\text{style loss}} + \underbrace{\gamma \mathcal{R}_{V\eta}}_{\text{TV regularizer}}$$

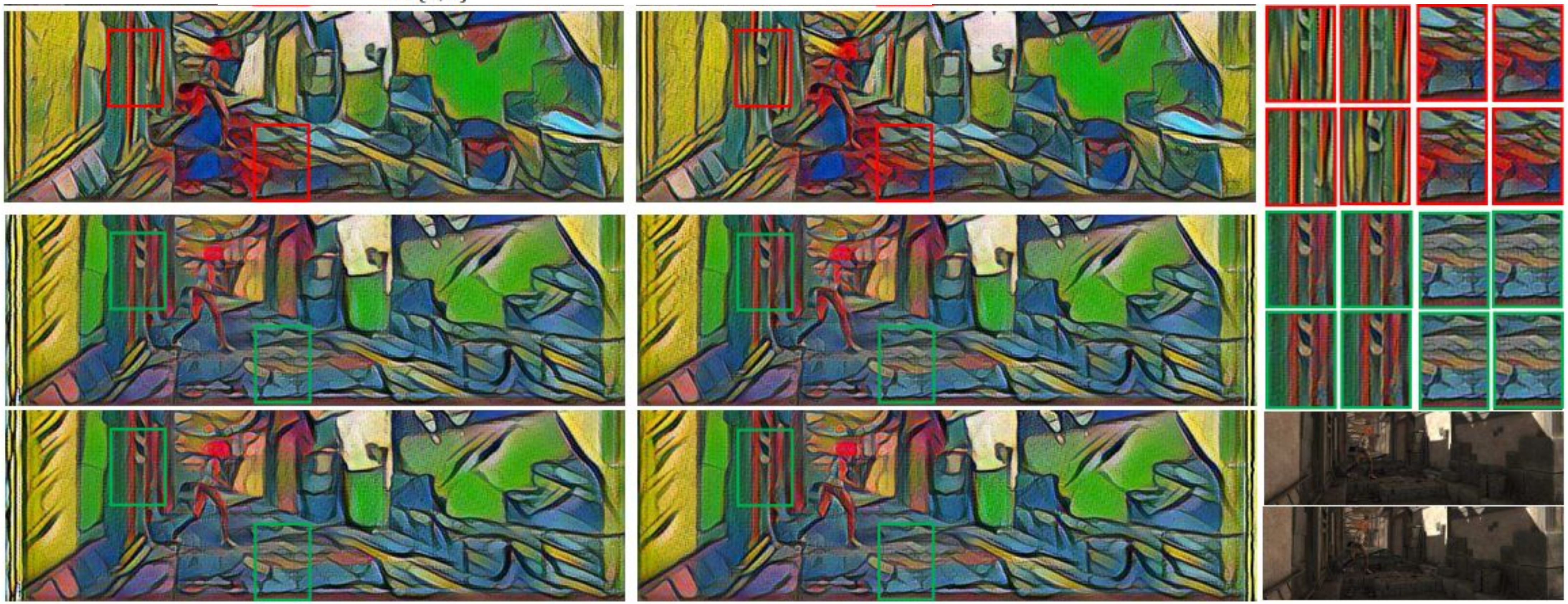


Real-Time Neural Style Transfer for Videos
Haozhi Huang
CVPR 2017

https://www.youtube.com/watch?v=BcflKNzO31A

Stereoscopic Neural Style Transfer

$$\mathcal{L}_{total} = \sum_{v\in\{l,r\}} \left(\alpha\mathcal{L}^{v}_{cont}(O_v, I_v) + \beta\mathcal{L}^{v}_{sty}(O_v, S) + \gamma\mathcal{L}^{v}_{disp}(O_v, D_v, M_v)\right).$$



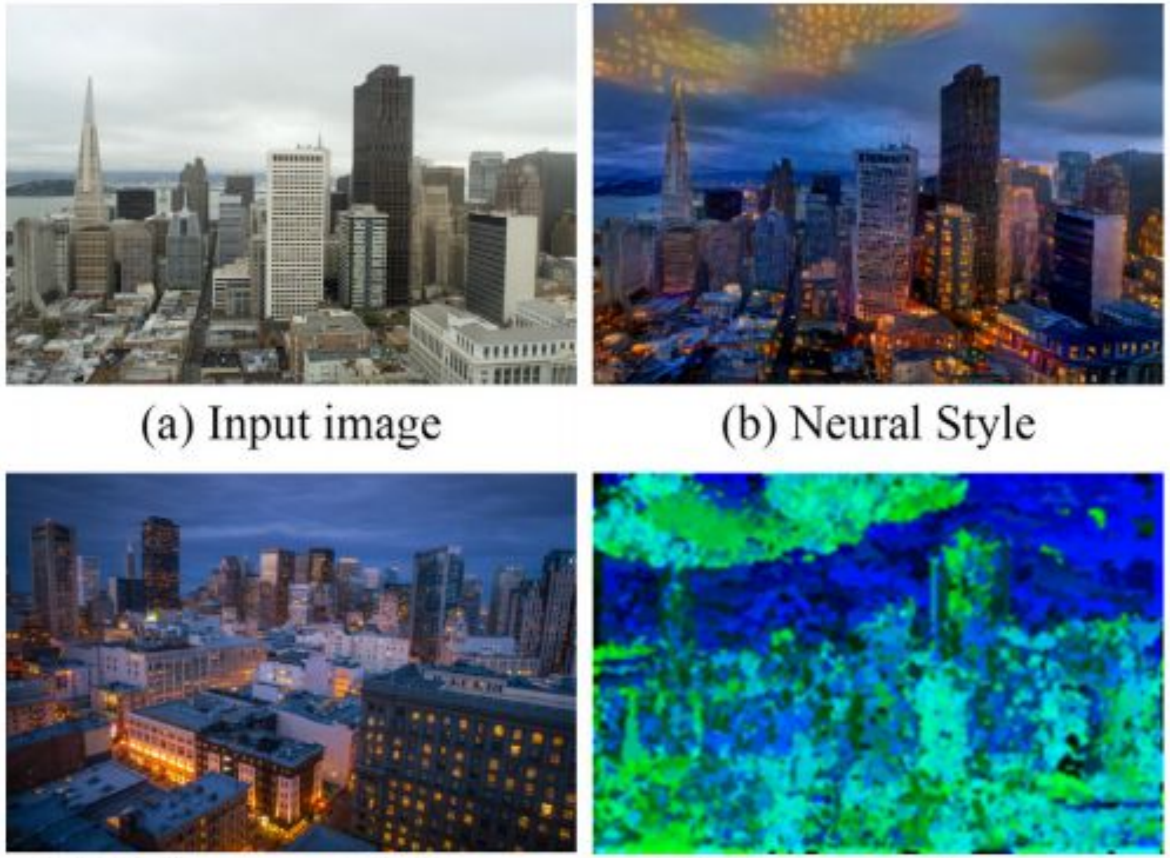
Stereoscopic Neural Style Transfer
Dongdong Chen et. al.,
CVPR 2018

(Photo Realistic) Style Transfer (1)

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_\ell \mathcal{L}_c^\ell + \Gamma \sum_{\ell=1}^{L} \beta_\ell \mathcal{L}_{s+}^\ell + \lambda \mathcal{L}_m$$

$$\mathcal{L}_m = \sum_{c=1}^{3} V_c[O]^T \mathcal{M}_I V_c[O]$$

- Transformation: locally affine in colorspace, and to express this constraint as a custom fully differentiable energy term.



(a) Input image (b) Neural Style (e) Reference style image (f) Correspondence of (b) and (e)

Deep Photo Style Transfer
Dongdong Chen et. al.,
CVPR 2017

(Photo Realistic) Style Transfer (1)

• Content

• Style

• Neural Style

• CNNMRF

• DPST

Deep Photo Style Transfer
Dongdong Chen et. al.,
CVPR 2017

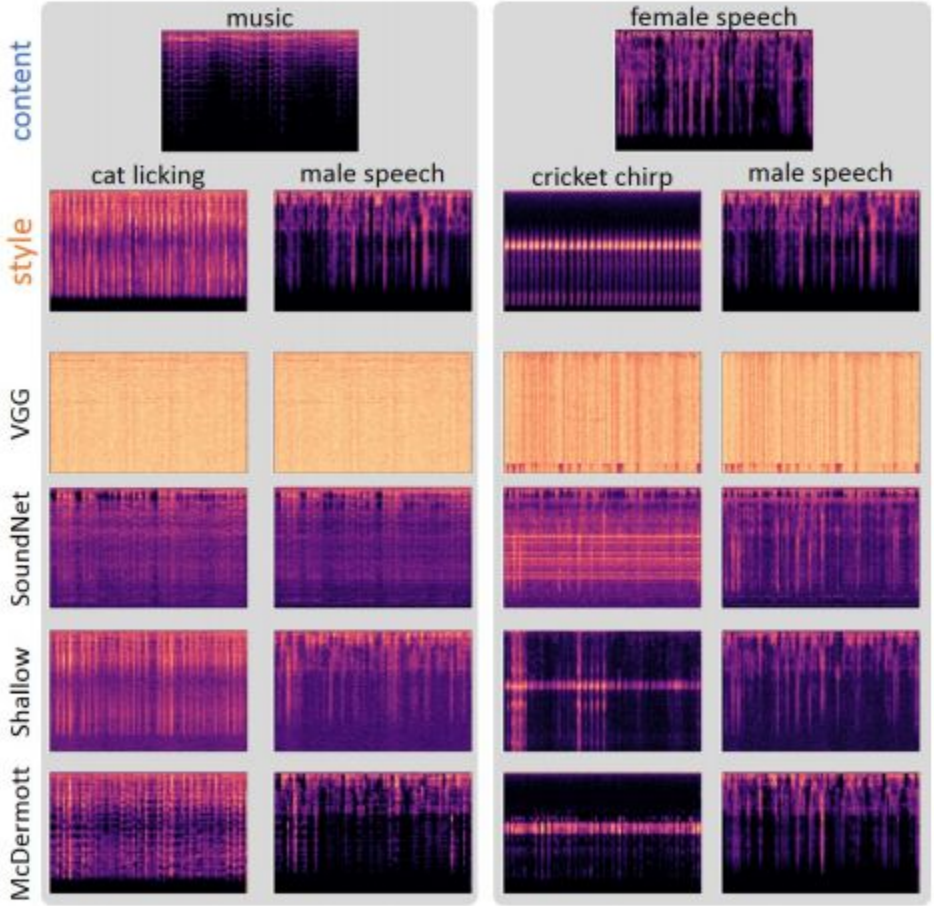
Language Style Transfer

Style Transfer Non-Parallel Text by Cross Alignment
Tianxiao Shen et. al.,
NIPS 2017

Model evaluation: 78.4% sentiment accuracy

Style Transfer for Audio and Music

Audio Style Transfer
Eric Grinstein, Ngoc Duong, Alexey Ozerov, Patrck Perez
ICASSP 2018



References

- Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps.K. Simonyan, A. Vedaldi, A. Zisserman. ICLR Workshop 2014
- Mahendran, Aravindh, and Andrea Vedaldi. "Understanding deep image representations by inverting them." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- Gatys, Leon, Alexander S. Ecker, and Matthias Bethge. "Texture synthesis using convolutional neural networks." Advances in Neural Information Processing Systems. 2015
- Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. "A neural algorithm of artistic style." *arXiv preprint arXiv:1508.06576* (2015).
- Li, Chuan, and Michael Wand. "Combining markov random fields and convolutional neural networks for image synthesis." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
- Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." European Conference on Computer Vision. Springer, Cham, 2016.
- Ulyanov, Dmitry, et al. "Texture Networks: Feed-forward Synthesis of Textures and Stylized Images." *ICML*. 2016.
- Ulyanov, D., A. Vedaldi, and V. S. Lempitsky. "Instance normalization: The missing ingredient for fast stylization. *ICML*. 2016.
- Huang, H., Wang, H., Luo, W., Ma, L., Jiang, W., Zhu, X., ... & Liu, W. (2017, July). Real-time neural style transfer for videos. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 7044-7052). IEEE.
- Ruder, Manuel, Alexey Dosovitskiy, and Thomas Brox. "Artistic style transfer for videos and spherical images." *International Journal of Computer Vision* (2018): 1-21.
- Chen, Dongdong, et al. "Stereoscopic neural style transfer." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Vol. 10. 2018.
- Luan, F., Paris, S., Shechtman, E., & Bala, K. (2017). Deep photo style transfer. *CVPR 2017*
- Shen, Tianxiao, et al. "Style transfer from non-parallel text by cross-alignment." *Advances in Neural Information Processing Systems*. 2017.
- Grinstein, E., Duong, N., Ozerov, A., & Perez, P. (2017). Audio style transfer. ICASSP 2018.