Computational Intelligence Laboratory

Significant Applications – Part I

Thomas Hofmann

ETH Zurich – cil.inf.ethz.ch

17 May 2019

Section 1

Image-to-Image Translation (pix2pix)

References

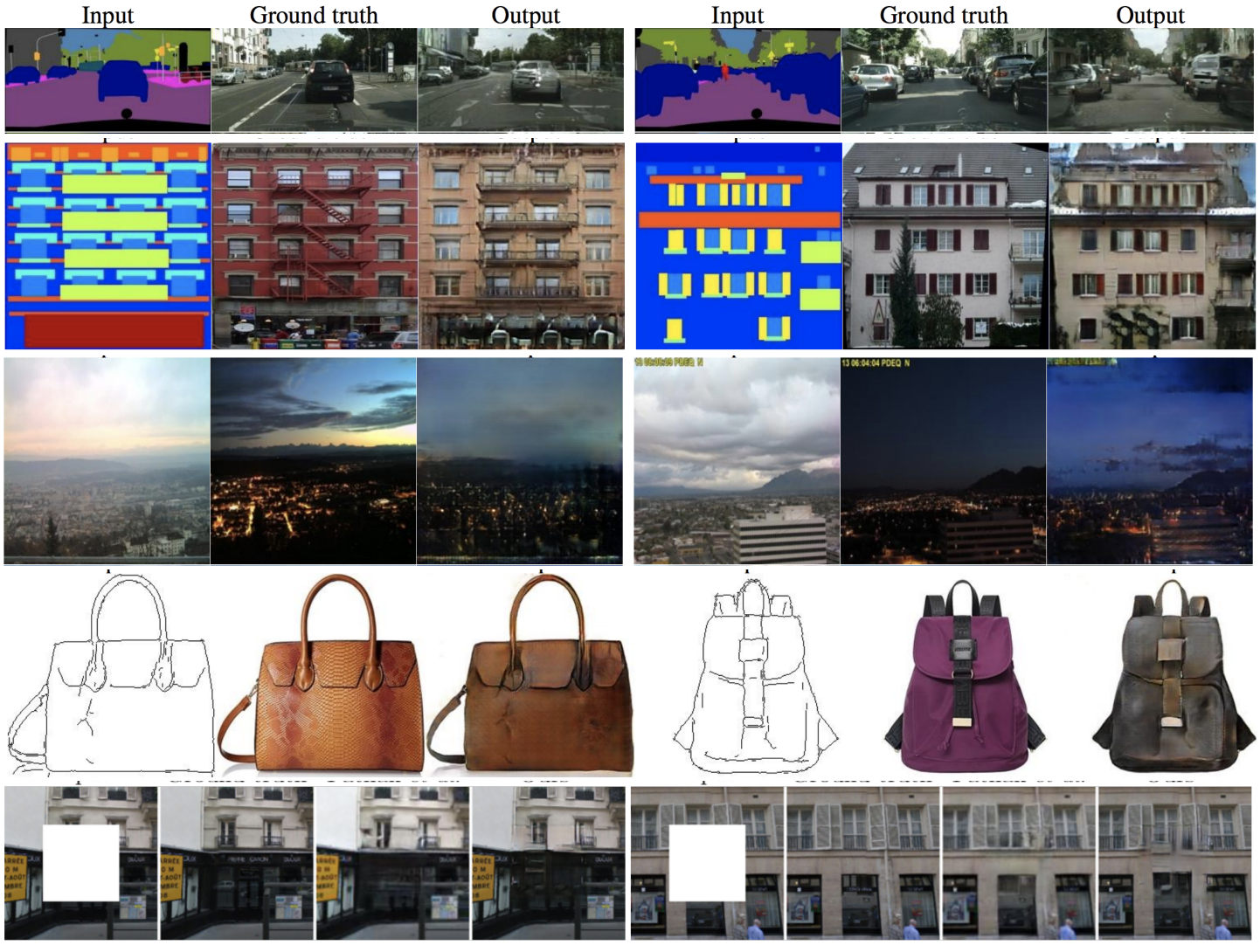
- Isola, P., Zhu, J. Y., Zhou, T., Efros, A. A.: Image-to-image translation with conditional adversarial networks. CVPR 2017.

- aka: **pix2pix** (2250 citations on Google scholar as of 2019-05-16)

Overview

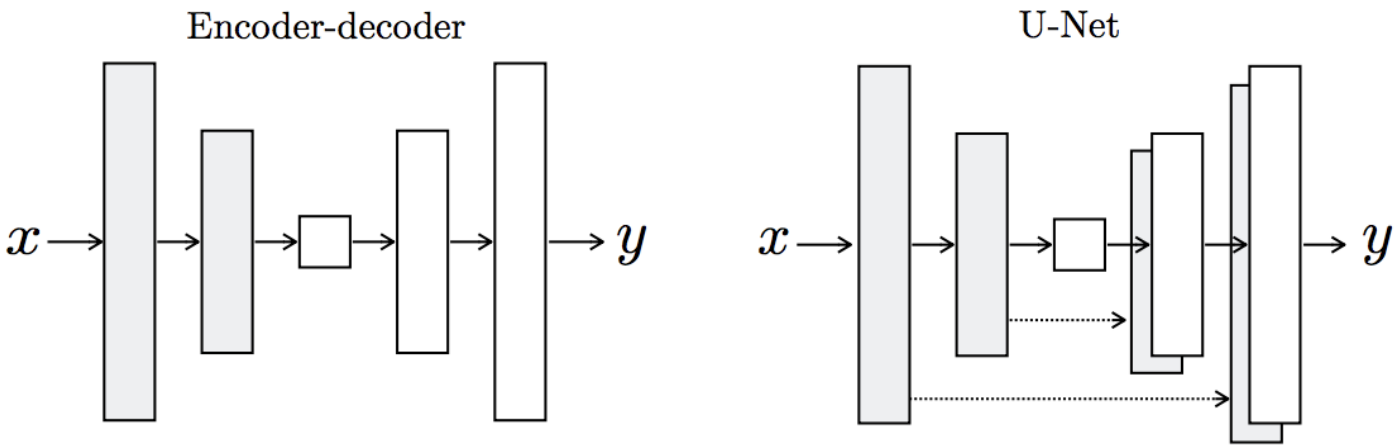
- Goal: learn $G : (\mathbf{y}, \mathbf{z}) \mapsto \mathbf{x}$, where $\mathbf{y}$ is the conditioned image, $\mathbf{z}$ is randomness and $\mathbf{x}$ is the generated image
- Investigate different design choices across a wide range of applications (and metrics)
  - Labels to Photos (more on this later)
  - Day to Night
  - Edges to Handbags / Shoes / Photo Models
  - Photo Impaiting
  - Thermal to Photos

Examples



Insights

- Unclear how to make use of noise (degenerates to determinstic)
- Combine GAN objective with L_1 distance $(++)$
- Condition GAN discriminator on input $(+++)$
- PatchGAN: discriminator looks at $\approx 70 \times 70$ pixel patches $(++)$
- Use U-net style skip connections $(+++)$



Demo

Interactive demo: `https://affinelayer.com/pixsrv/`

Video: Pix2Pix Unity: `https://vimeo.com/287778343`

Section 2

Photorealistic Image Generation

References

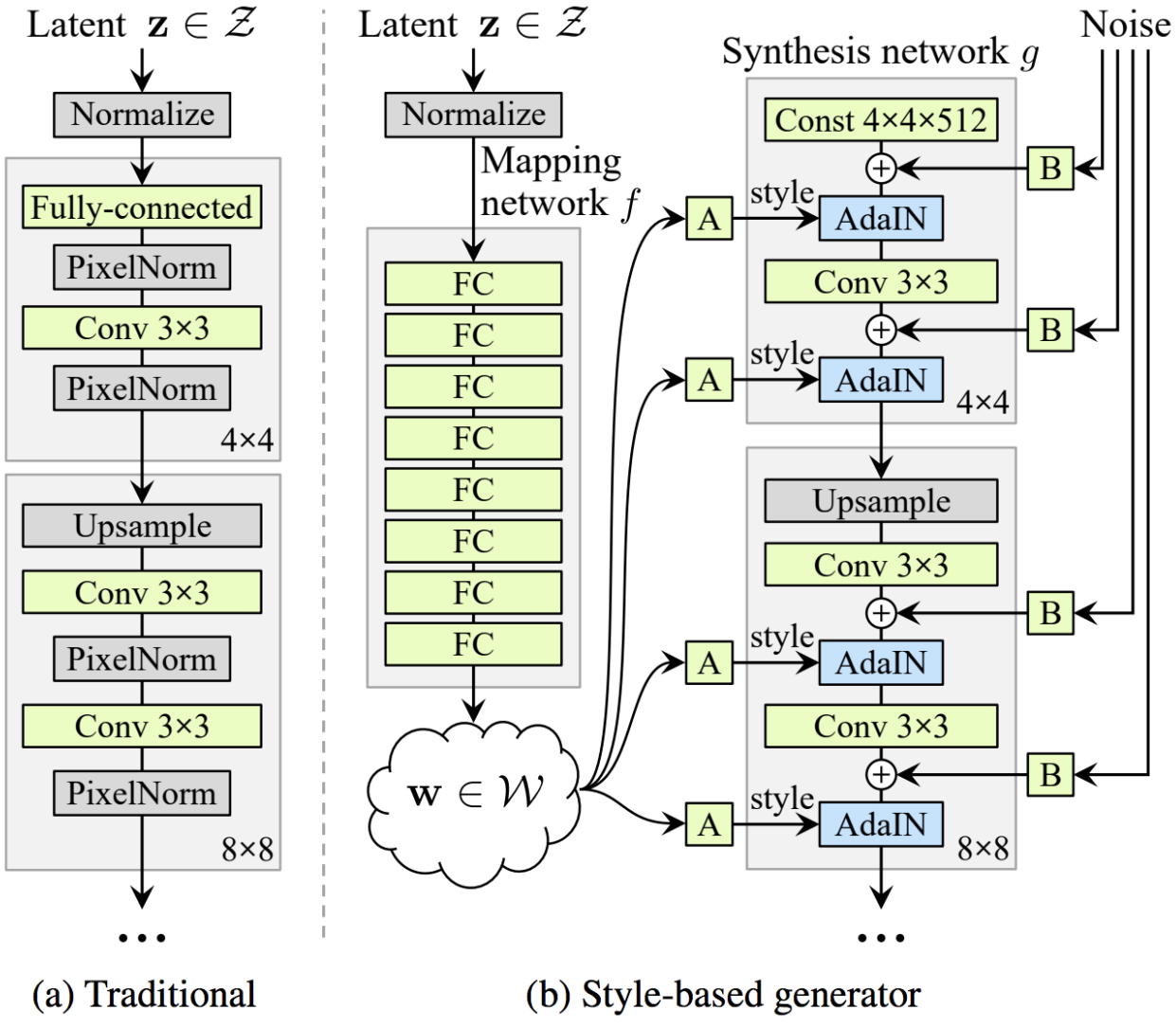
- Karras, T., Laine, S., Aila, T.: A Style-Based Generator Architecture for Generative Adversarial Networks, CVPR 2019. [pdf]

- Video: `https://www.youtube.com/watch?v=kSLJriaOumA` (6'17")

Overview

- Objectives:
  - High quality photorealistic image generation
  - Controlled process of image synthesis: style + stochasticity
  - Disentanglement of modes of variation
- Ideas:
  - GAN, multiscale image generation
  - Split randomness between style control and stochastic noise
  - Novel use of latent code: transformation, self-modulation

Architecture



(a) Traditional

(b) Style-based generator

Multiscale Generation

- Start with small resolution: 4x4 (constant input, 512 channels)
- Successively up-sample image to 8x8, . . . , 1024x1024
- Repeated processing module between scales
  - 2x: 3x3 convolution, add noise, style transform
- Previous Work
  - Laplacian Pyramid + GAN: Denton&, NIPS 2015
  - DCGAN: Radford&, ICLR 2016

Transformation & Injection of Randomness

- Style Randomness
  - Non-linear transformation of simple noise vector (fully connected DNN, 512 to 512 dimensions, **mapping network**)
  - Learned **affine transformation** at each injection point
  - **Adaptive instance normalization**

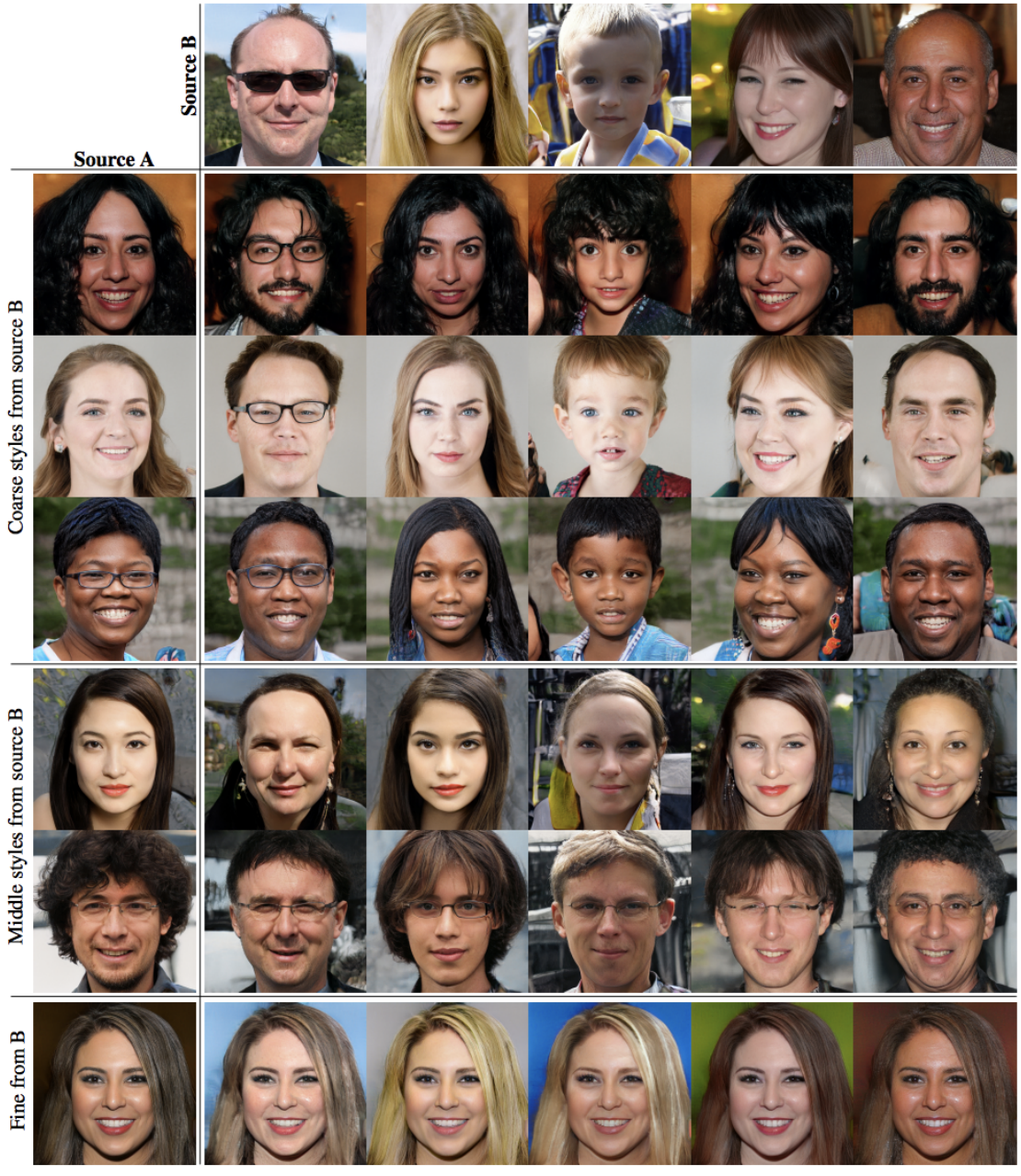
$$F_{r,c}(\mathbf{x}, \mathbf{y}) = y^s_{r,c} \frac{x_{r,c} - \mu(\mathbf{x})}{\sigma(\mathbf{x})} + y^b_{r,c}$$

    - μ, σ: averaged spatially per channel and per instance
- Noise
  - Gaussian white noise images (per layer)
  - learned gain (per channel)

Transformation & Injection of Randomness

- Previous Work
  - Instance normalization [...]: Ulyanov&, arXiv 2016
  - Arbitrary style transfer in real-time with adaptive instance normalization: Xun & Belongie, CVPR 2017
  - A learned representation for artistic style: Dumoulin&, ICLR, 2017

Results


Source B
Source A
Coarse styles from source B
Middle styles from source B
Fine from B

Section 3

Semantically-Conditioned Image Synthesis

References

- Park, T., Liu, M. Y., Wang, T. C., Zhu, J. Y.: Semantic Image Synthesis with Spatially-Adaptive Normalization. CVPR 2019. [github]

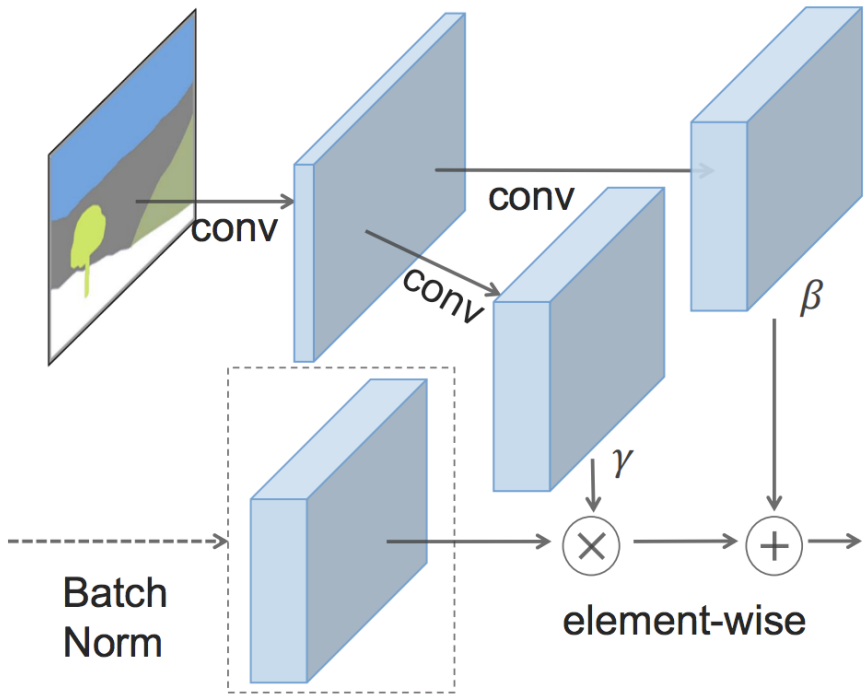
SPADE

- Goal: synthesize photorealistic images given a **semantic layout**.

- Approach: Use semantic segmentation to modulate activations in normalization layers through a spatially adaptive, learned transformation.

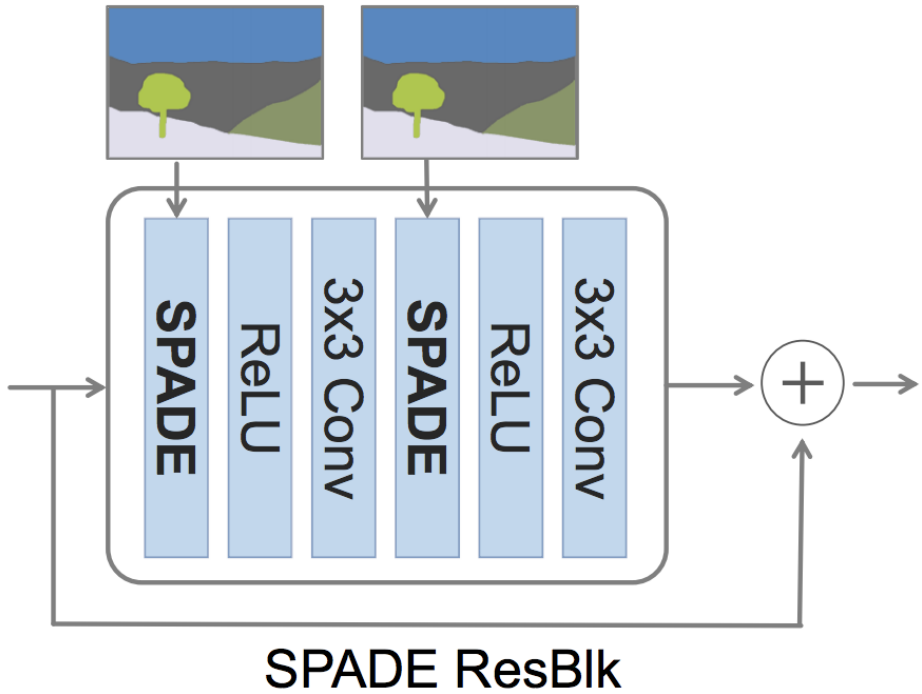
SPADE Layers

- Normalized activations are denormalized by modulating the activation using a learned affine transformation whose parameters are inferred from external data



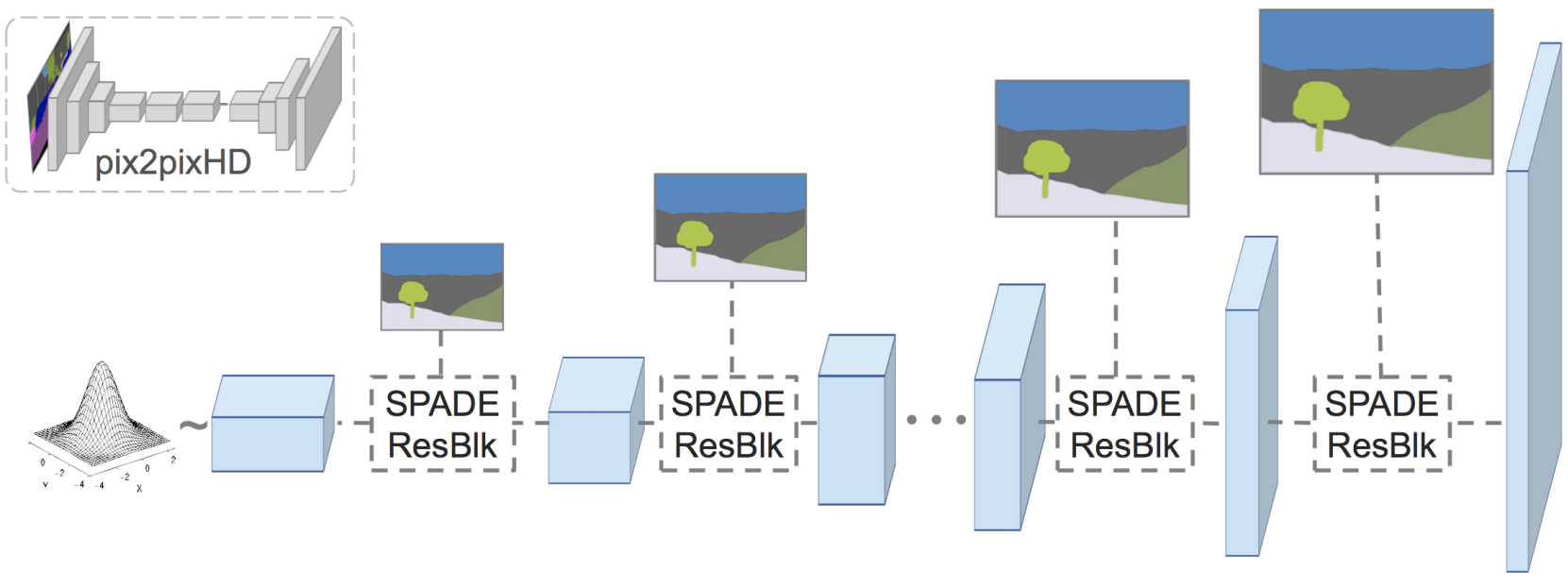
SPADE Architecture

- SPADE residual layer block

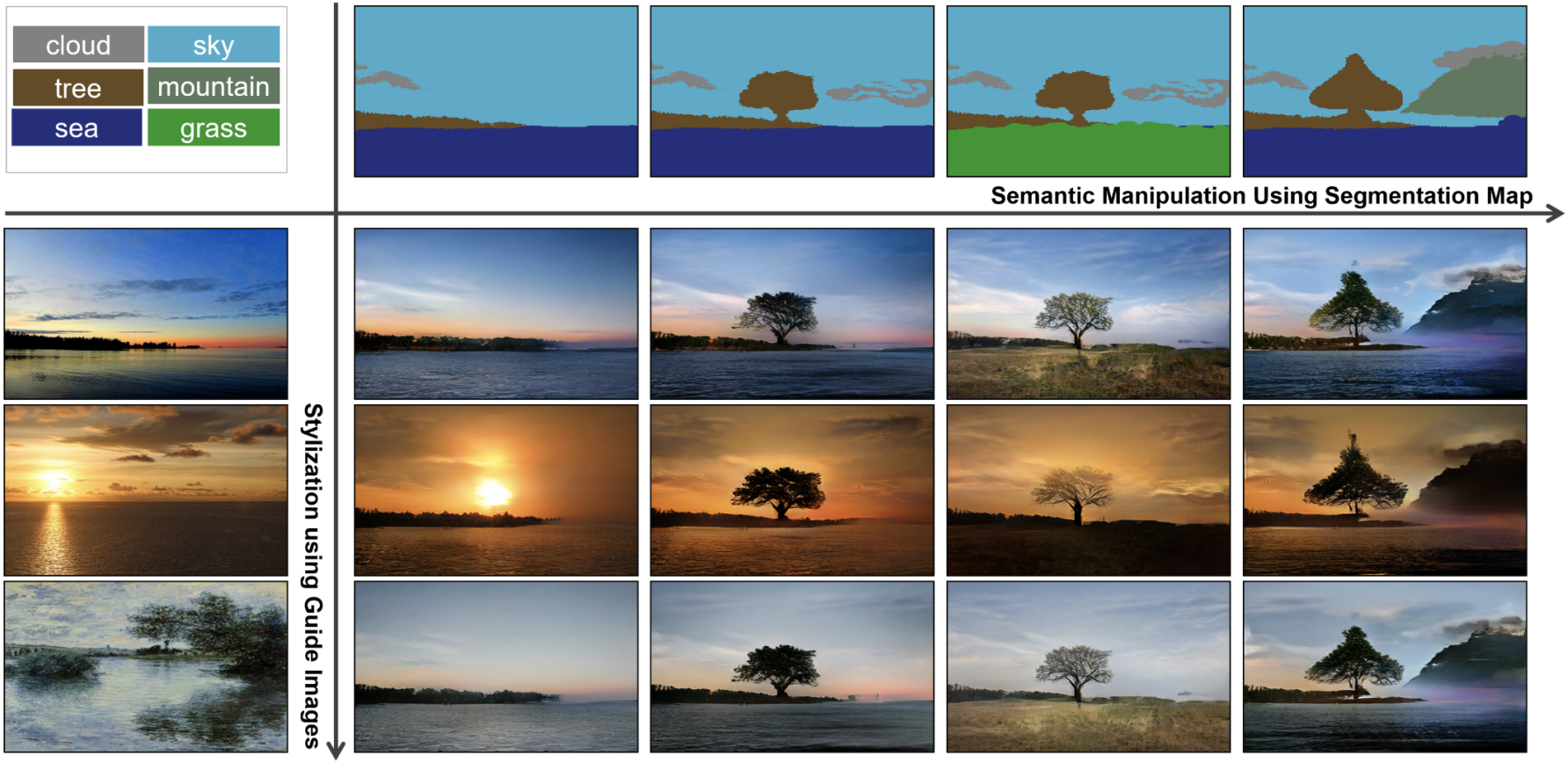


SPADE Architecture

- SPADE overall architecture



Results


cloud
sky
tree
mountain
sea
grass
Semantic Manipulation Using Segmentation Map
Stylization using Guide Images

Results

Video: https://youtu.be/MXWm6w4E5q0 (2'17")

Section 4

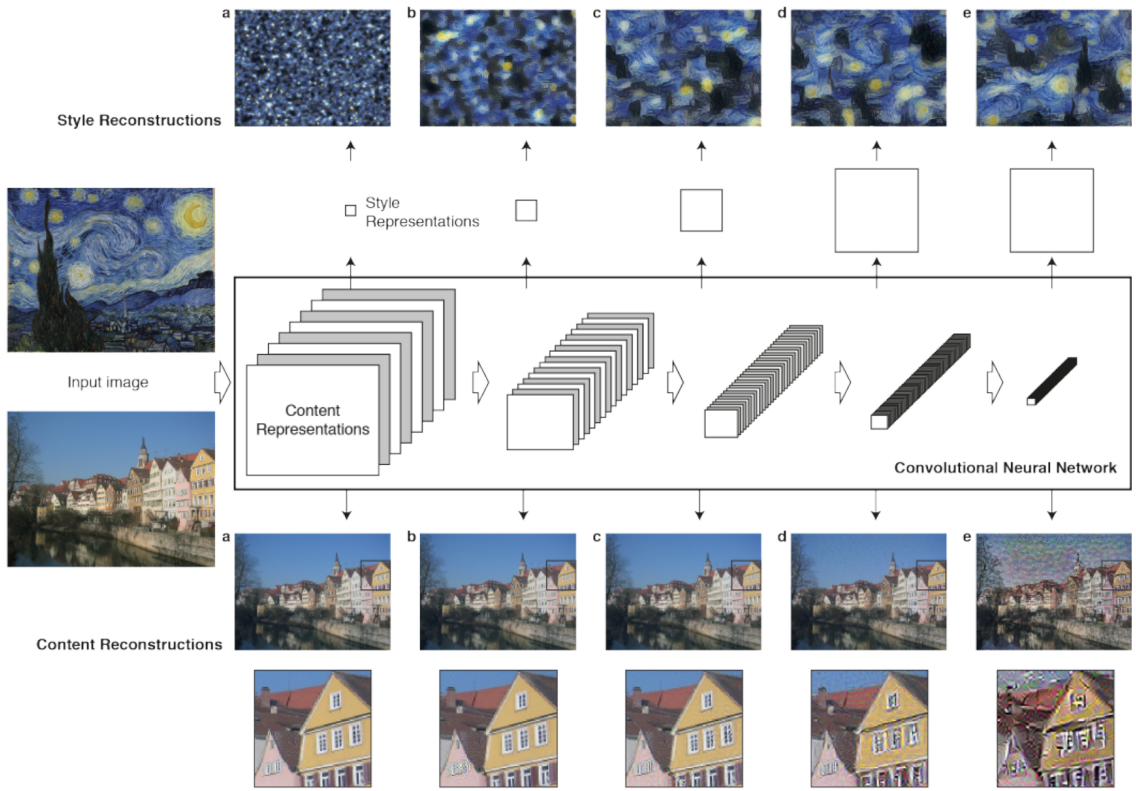
Artistic Image Generation

References

- Gatys, L. A., Ecker, A. S., Bethge, M.: A Neural Algorithm of Artistic Style, Journal of Vision 2015
- Ruder, M., Dosovitskiy, A., Brox, T.: Artistic style transfer for videos. GCPR 2016. [video]

Content and Style Representations

- **Content network**: VGG network trained for image recognition (15 conv, 4 pooling layers)
- **Style network**: Fixed, computes pairwise correlations between feature channels



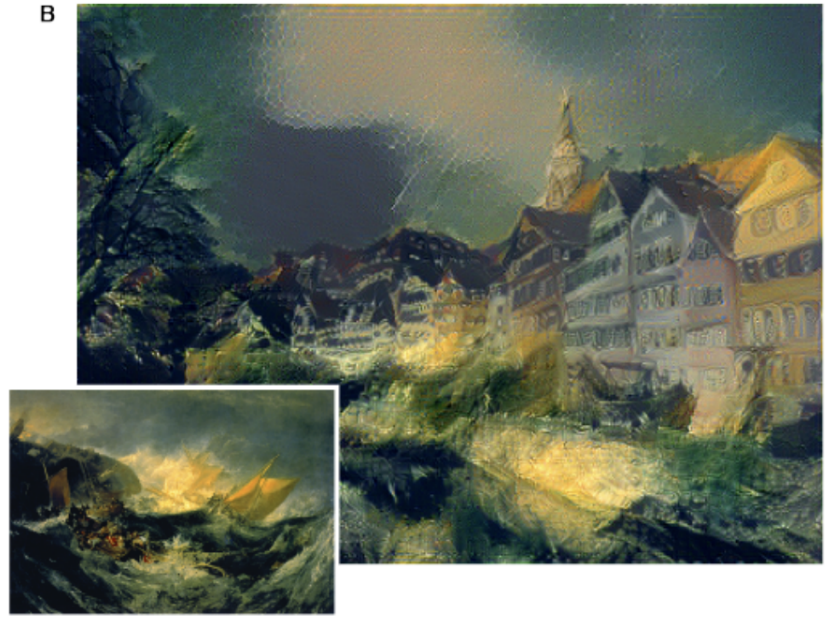
Reconstruction from Feature Maps

- Implicit inversion (approximate, non-unique): from features back to input $\mathbf{x}$
  - Start with random input $\mathbf{x} = \mathbf{x}^0$
  - Loss function: squared feature loss between given features and observed features $F(\mathbf{x})$
  - Perform gradient descent w.r.t. inputs (not weights) and update input $\mathbf{x}$
  - Can be done for content as well as style feature maps (see figure above)
- Mixing of content and style
  - Find images that matches content features of the original (e.g. at specific layer) ... and at the same time ....
  - whos style features map the style features of the style example image

Results

A

B

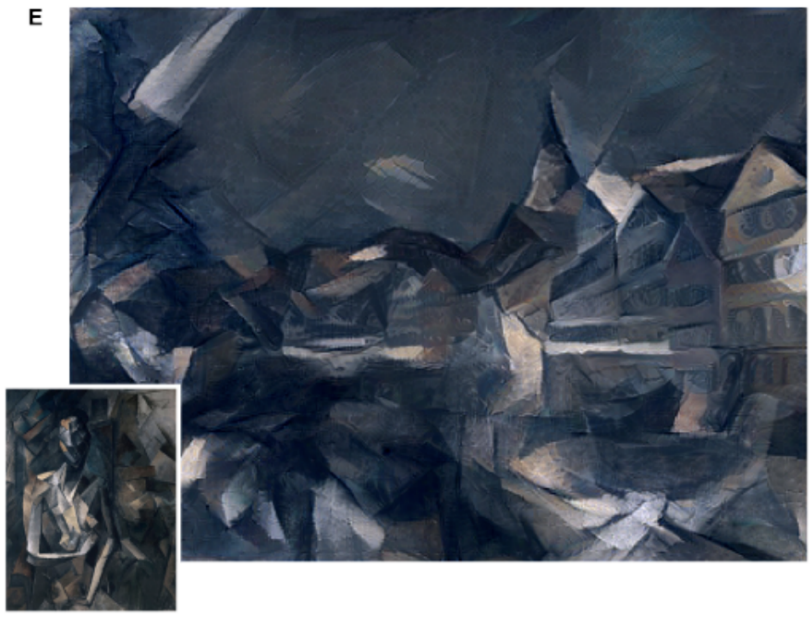


Results

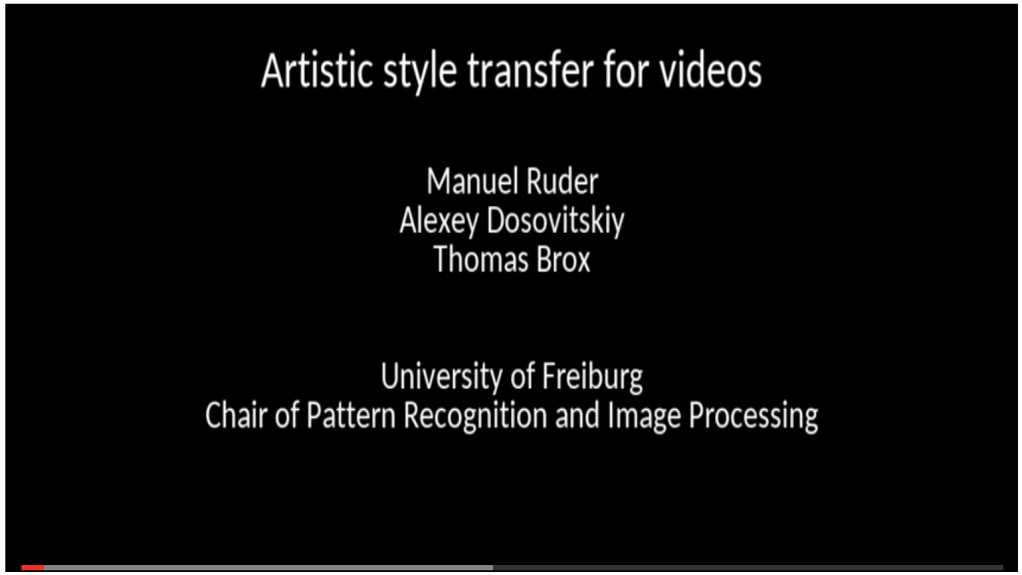
C

D

Results



Videos


Artistic style transfer for videos
Manuel Ruder
Alexey Dosovitskiy
Thomas Brox
University of Freiburg
Chair of Pattern Recognition and Image Processing

Section 5

Towards Higher Image Fidelity

Trading-off Fidelity with Diversity

Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. arXiv:1809.11096, 2018.