Deep Feature Interpolation for Image Content Changes

by P. Upchurch, J. Gardner, G. Pleiss, R. Pleiss, N. Snavely, K. Bala, K. Weinberger

Presenter: Aliya Amirzhanova

Master Scientific Computing,

Heidelberg University

Date: 28.06.2018

Content

- Introduction
  - Why are image content transformations inspirational?
  - Related work
  - Motivation

- Deep feature interpolation
  - Algorithm
  - Details of the procedure

- Experimental results
  - Dataset
  - Visual results

- Conclusion
  - Limitations
  - Summary
  - Future work

Why are image content transformations inspirational?

"When the image is new, the world is new."
Gaston Bachelard, The Poetics of Space

Why are image content transformations inspirational?

Fig.1 Editing image content [1]

Why are image content transformations inspirational?

- Active and challenging research area in computer vision and graphics
- Semantic transformation -> answers to questions: ‘What if…?’
- Realistic results
- Applications
- Image inpainting

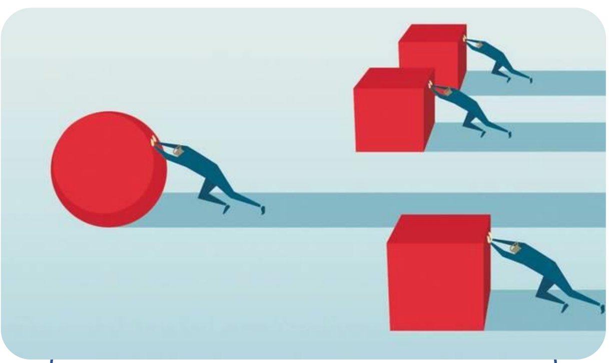
Related work

- Generative methods: Larsen et al. (2015) [2] and Radford et al. (2016) [3]
- Manipulation of latent space variables [4]
- Minimizing the witness function of MMD: Gardner et al. (2015) [5]
- Optimization of feature targets during reconstruction: Gatys et al. (2015) [6]
- Combination of several vision methods [7]

Motivation

Simple

Versatile



Fast

Deep Feature Interpolation

Notations:

x – test image

z – output image

$S^t = \{x_1^t, \dots, x_n^t\}$ – a set of target images with the desired attribute

$S^s = \{x_1^s, \dots, x_m^s\}$ – a set of source images without the attribute

x -> *φ(x)** – obtaining a new representation of an image

φ(x) – a vector consisting of concatenated activations of the convnet* when applied to image *x*

*Provided with pre-trained VGG network

Algorithm:

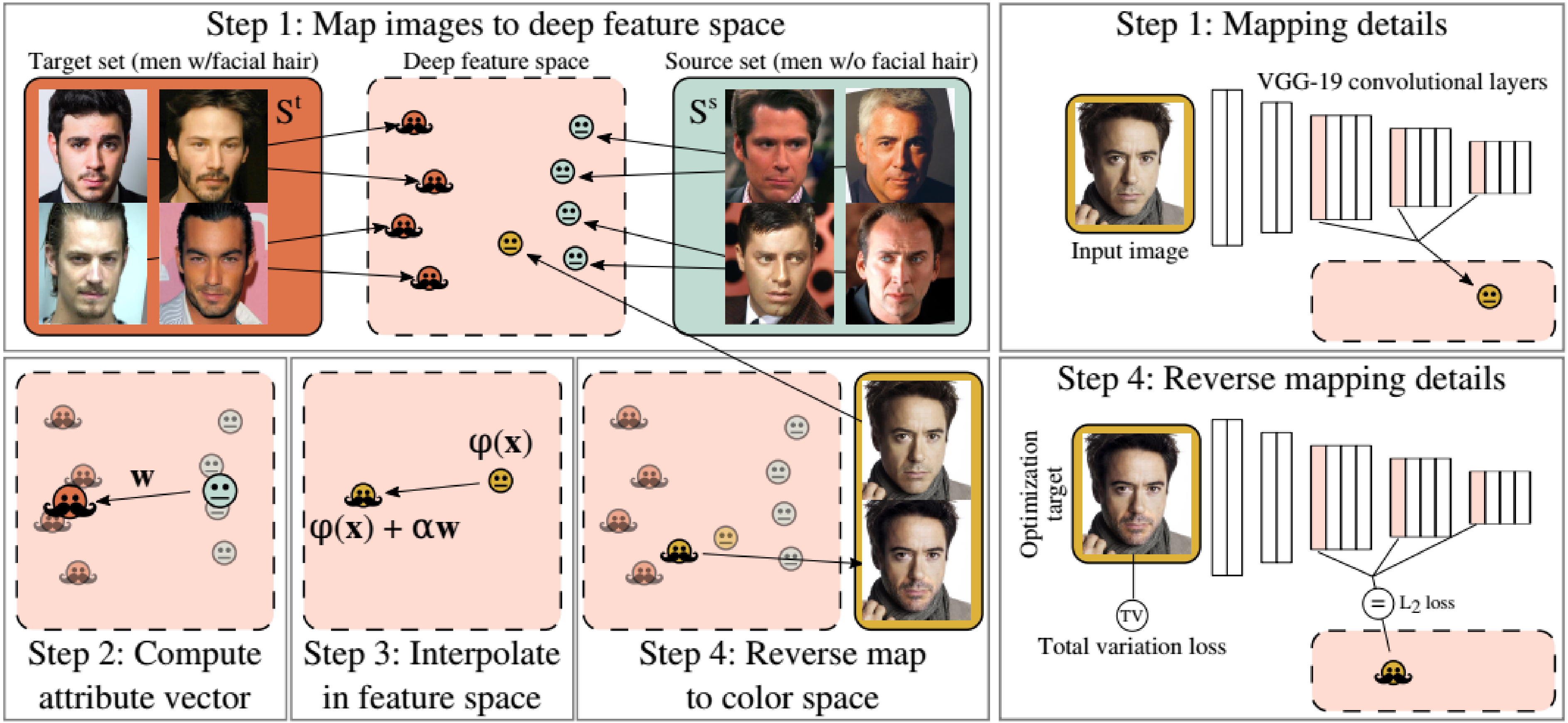


Fig.2 A schematic outline of DFI [1]

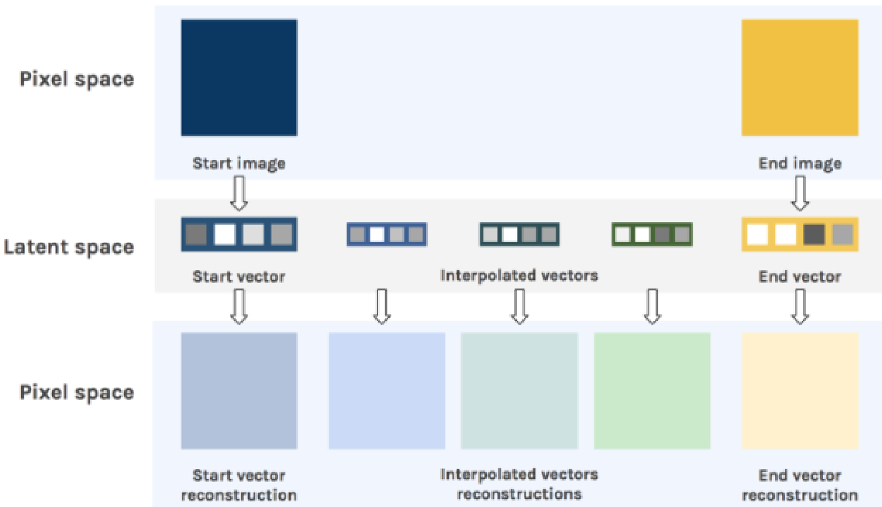
Details of the procedure:

- Selecting S^t and S^s
  - Assumption: the attribute vector w isolates the targeted information
    - *w = φ(z) - φ(x)*
  - Both sets should be as similar as possible
  - To ensure sufficient similarity, restrict to the K nearest neighbors of S^t to *φ(x)*
    - $\overline{\varphi}^t = \frac{1}{K}\sum_{x^t \epsilon N_K^t} \varphi(x^t)$ and $\overline{\varphi}^s = \frac{1}{K}\sum_{x^s \epsilon N_K^s} \varphi(x^s)$
- Deep feature mapping
  - The deep feature space should be suitable for:
    - linear interpolation
    - reverse mapping back into pixel space

Details of the procedure:

- Image transformation
  - $\overline{\varphi}^t$ and $\overline{\varphi}^s$ will have very small components in most features
  - Features unrelated to the target attribute will be averaged to very small values and approximately subtracted away in the vector *w*
- Reverse mapping
  - z = $argmin_z \frac{1}{2} \|(\varphi(x) + \alpha w) - \varphi(z)\|_2^2 + \lambda_{V^\beta} R_{V^\beta}(z)$
  - R_{V^β} - total variation regularizer
  - $R_{V^\beta}(z)$ = $\sum_{i,j}((z_{i,j+1} - z_{i,j})^2 + (z_{i+1,j} - z_{i,j})^2)^{\frac{\beta}{2}}$
  - λ_{V^β} = 0.001 and β = 2

Algorithm:



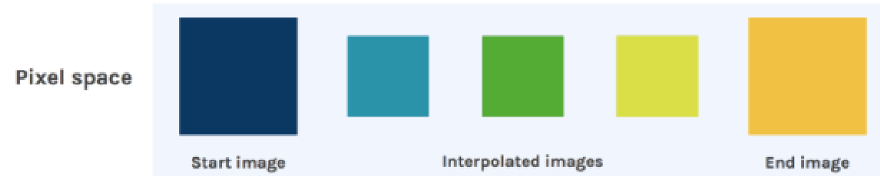


Fig.3 Interpolation in latent space [8]

Experimental results

Dataset

- Labeled Faces in the Wild (LFW)
    - 13, 143 images of faces
    - Predicted annotations for 73 different attributes
    - Six attributes for testing
        - senior, mouth open, eyes open, smiling, moustache, eyeglasses
- A high resolution dataset from CelebA, MegaFace and Helen + Google image search
- A shoes subset of UT Zappos50k


Older
Mouth Open
Eyes Open
Smiling
Moustache
Glasses
Older
Mouth Open
Eyes Open
Smiling
Moustache
Glasses
Original
Deep Feature Interpolation
AEGAN

Experimental results

older	mouth open	eyes open	smiling	moustache	glasses
4.57	7.09	17.6	20.6	24.5	38.3

Table.1 Perceptual study results [1]

Experimental results

Fig.4 Editing high resolution images [1]

Experimental results

Fig.5 Inpainting. LFW faces [1]

Experimental results

Fig.6 Inpainting. UT Zappos50k shoes [1]

Varying free parameters

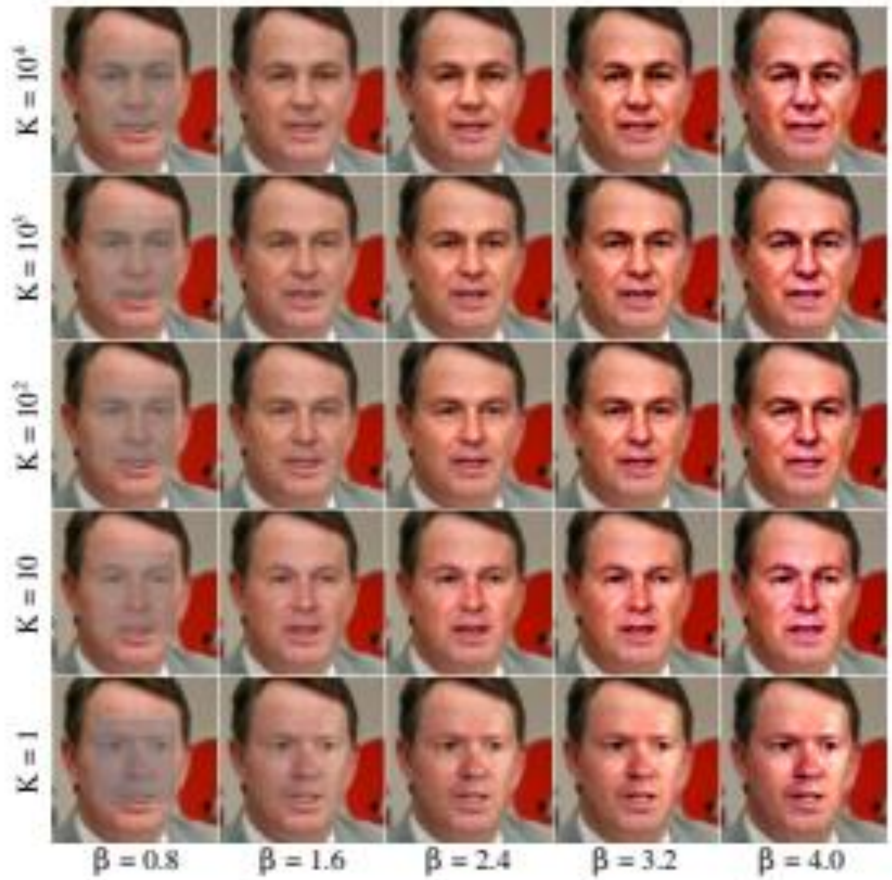


Fig.7 The effect of changing β and K [1]

Varying free parameters

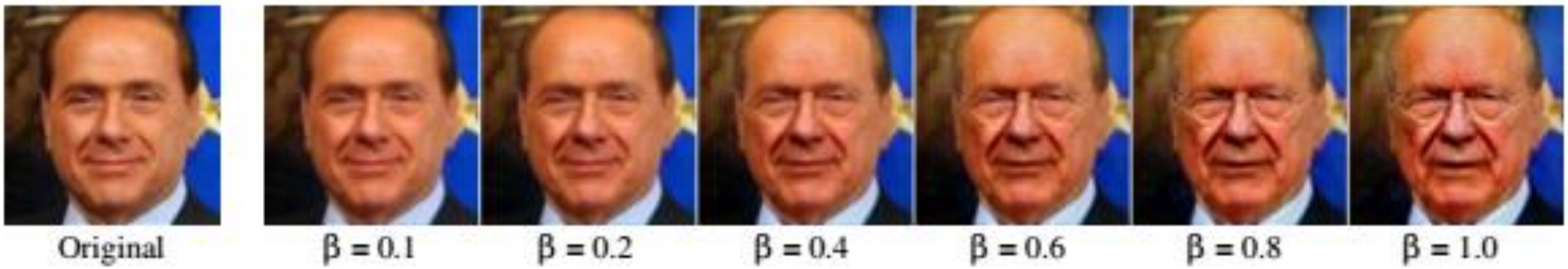


Fig.8 The effect of changing β [1]

Limitations

- Image alignment
- DFI is incapable of shape or rotation transformations
- Sample images have to be similar to target image
- DFI is unable to reconstruct the image properly when the masked region is a half of the image
- DFI is not powerful enough for complex tasks

Fig.9 DFI results of an image with the right half missing [1]

Summary

- Wide range image transformations
- Can be used as a baseline method
- According to the authors DFI is the first algorithm to enable automated high resolution content transformations

Future work

- Possibility to incorporate techniques from real-time style transfer to speed-up DFI

Thank you for your attention!
Questions?

References

[1] P. Upchurch, J. Gardner, G. Pleiss, R. Pleiss, N. Snavely, K. Bala, K. Weinberger (2017) Deep Feature Interpolation for Image Content Change. https://arxiv.org/abs/1611.05507

[2] A.B.L. Larsen, S.K. Sonderby, H. Larochelle, O. Winther (2015) Autoencoding beyond pixels using a learned similarity metric. https://arxiv.org/abs/1512.09300

[3] A. Radford, L. Metz, S.Chintala (2016) Unsupervised representation learning with deep convolutional generative adversarial networks.

[4] A. Brock, T. Lim, J. Ritchie, N. Weston (2016) Neural photo editing with introspective adversarial networks. https://arxiv.org/abs/1609.07093

[5] J.R. Gardner, P. Upchurch, M.J.Kusner, Y.Li, K. Weinberger, K. Bala, J.E. Hopcroft (2015) Deep manifold traversal: changing labels with convolutional features. https://arxiv.org/abs/1511.06421

[6] L.A. Gatys, A.S. Ecker, M. Bethge (2015) A neural algorithm of artistic style. https://arxiv.org/abs/1508.06576

References

[7] S. Suwajanakorn, S.M. Seitz, I. Kemelmacher-Shlizerman (2015) What makes Tom Hanks look like Tom Hanks.

[8] https://hackernoon.com/latent-space-visualization-deep-learning-bits-2-bd09a46920df

[9] https://memegenerator.net/instance/72296984/waiting-skeleton-near-computer-still-waiting-for-my-neural-network-to-train

[10] http://www.juliehyde.com.au/resisting-the-pull/

[11] https://asegrad.tufts.edu/academics/explore-graduate-programs/diversity-and-inclusion-leadership

[12] https://englishhelponline.me/2011/07/25/the-difference-between-words-fast-and-early/

[13] http://www.thehindu.com/opinion/lead/learning-to-run-twice-as-fast/article18260719.ece

[14] https://www.iconspng.com/image/52040/tortoise-and-hare-fast-and-slow