Pixel-based Facial Expression Synthesis

Arbish Akram and Nazar Khan

Computer Vision and Machine Learning Group,
Punjab University College of Information Technology (PUCIT)
University of the Punjab, Lahore, Pakistan

25th International Conference on Pattern Recognition
January 14, 2021

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Introduction

- Facial expression synthesis (FES) has achieved remarkable advances with the advent of Generative Adversarial Networks (GANs).

Introduction

- Facial expression synthesis (FES) has achieved remarkable advances with the advent of Generative Adversarial Networks (GANs).
- GAN-based FES models limitations:
    1. Generate photo-realistic results as long as testing images are similar to training images.

Introduction

- Facial expression synthesis (FES) has achieved remarkable advances with the advent of Generative Adversarial Networks (GANs).
- GAN-based FES models limitations:
  1. Generate photo-realistic results as long as testing images are similar to training images.
  2. Require thousands of images for training.

Introduction

- Facial expression synthesis (FES) has achieved remarkable advances with the advent of Generative Adversarial Networks (GANs).
- GAN-based FES models limitations:
  1. Generate photo-realistic results as long as testing images are similar to training images.
  2. Require thousands of images for training.
  3. Higher computational and storage resources at testing time.

Motivation

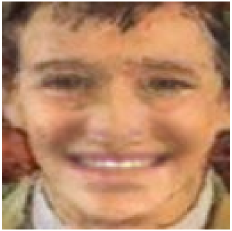
Recently, Masked Regression (MR)[1] has shown that

- facial expressions usually constitute local instead of global changes
- transformation from neutral to happy mostly affects the regions around the eyes, nose and mouth to induce happy expression.

[1] Nazar Khan et al. (2020). "Masked Linear Regression for Learning Local Receptive Fields for Facial Expression Synthesis." In: *International Journal of Computer Vision* 128.5, pp. 1433–1454.

Motivation by MR

We propose a regression-based method that looks at only one fixed input pixel to produce an output pixel.



Contributions

- First pixel-based ridge regression method to solve the FES problem.

Contributions

- First pixel-based ridge regression method to solve the FES problem.
- Pixel-based idea can be extended using kernel regression.

Contributions

- First pixel-based ridge regression method to solve the FES problem.
- Pixel-based idea can be extended using kernel regression.
- The proposed method
    1. generalizes much better for a variety of out-of-dataset images.

Contributions

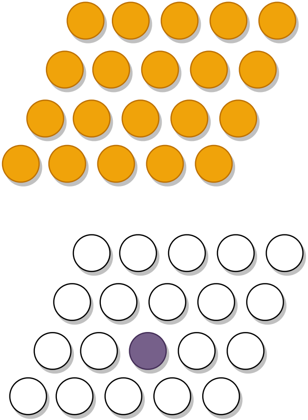
- First pixel-based ridge regression method to solve the FES problem.
- Pixel-based idea can be extended using kernel regression.
- The proposed method
  1. generalizes much better for a variety of out-of-dataset images.
  2. is two orders of magnitude smaller than GAN-based models.

Contributions

- First pixel-based ridge regression method to solve the FES problem.
- Pixel-based idea can be extended using kernel regression.
- The proposed method
  1. generalizes much better for a variety of out-of-dataset images.
  2. is two orders of magnitude smaller than GAN-based models.
  3. can be deployed in mobile devices and embedded systems.

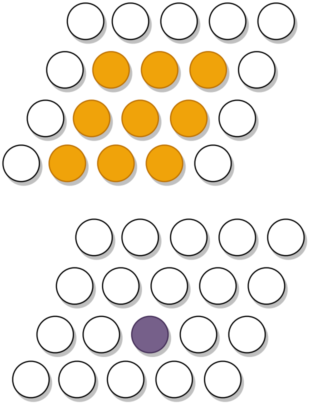
Ridge Regression (RR)

- The output of p^{th} pixel is produced by looking at all input pixels.



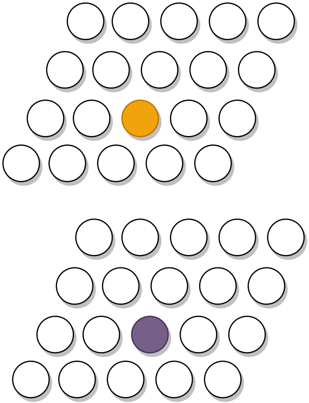
Masked Regression (MR)

- The output of p^{th} pixel is produced by looking at a local patch of input pixels.



Pixel-based Ridge Regression (Pixel-RR)

- The output of p^{th} pixel is produced by looking at only one input pixel.



Pixel-RR

- Objective function for Pixel-RR

$$E(w_p, b_p) = \frac{1}{2}\|w_p\mathbf{x}_p + b_p\mathbf{1} - \mathbf{t}_p\|_2^2 + \frac{\lambda}{2}(w_p^2 + b_p^2)$$

- Scalars w_p and b_p are learnable weight and bias values.
- $\mathbf{x}_p$ and $\mathbf{t}_p \in \mathcal{R}^{1\times N}$.
- Unique global minimizers can be computed as

$$\begin{bmatrix} w_p \\ b_p \end{bmatrix} = \begin{bmatrix} \mathbf{x}_p\mathbf{x}_p^T + \lambda & \mathbf{1}\mathbf{x}_p^T \\ \mathbf{1}\mathbf{x}_p^T & N + \lambda \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{t}_p\mathbf{x}_p^T \\ \mathbf{t}_p\mathbf{1}^T \end{bmatrix}$$

Pixel-based Kernel Regression (Pixel-KR)

- Pixel-based mapping idea can be extended by using kernel regression.

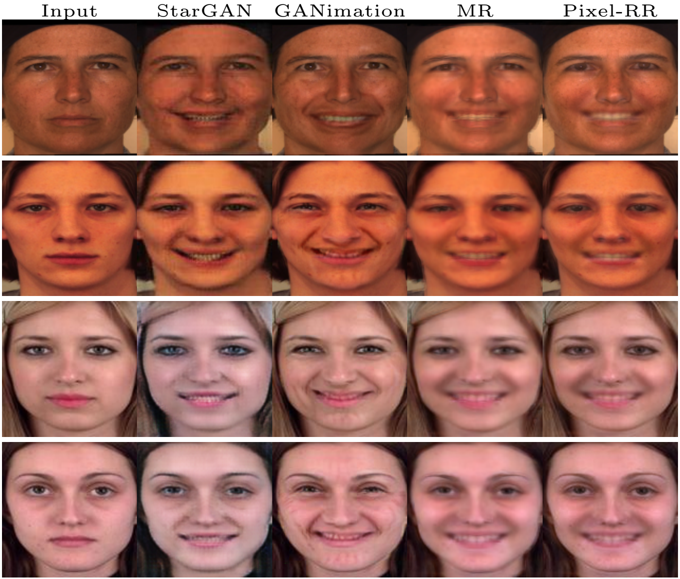
$$
\begin{aligned}
E(\mathbf{c}_p) &= \frac{1}{2}\|\mathbf{c}_p\phi(\mathbf{x}_p)^T\phi(\mathbf{x}_p) - \mathbf{t}_p\|_2^2 + \frac{\lambda}{2}\|\mathbf{c}_p\phi(\mathbf{x}_p)^T\|_2^2 \\
&= \frac{1}{2}\|\mathbf{c}_pK_p - \mathbf{t}_p\|_2^2 + \frac{\lambda}{2}\mathbf{c}_pK_p\mathbf{c}_p^T
\end{aligned}
$$

- $K_p = \phi(\mathbf{x}_p)^T\phi(\mathbf{x}_p) \in \mathcal{R}^{N\times N}$ is the kernel matrix.
- Projection matrix $\mathbf{c}_p \in \mathcal{R}^{1\times N}$ can be computed as:

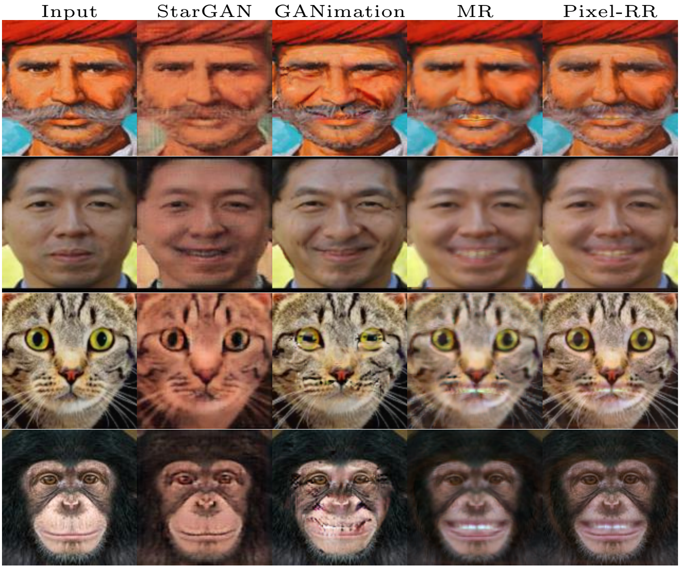
$$
\mathbf{c}_p = \mathbf{t}_p(K_p + \lambda I)^{-1}
$$

Results

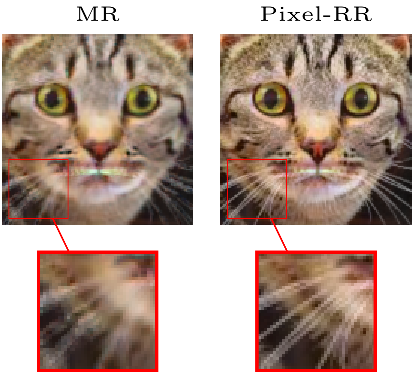
Qualitative results on in-dataset images



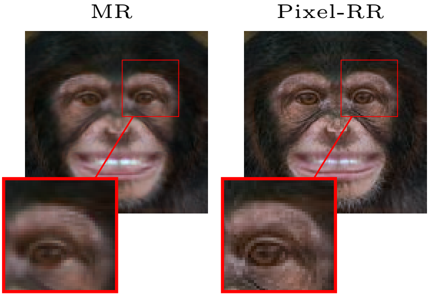
Qualitative results on out-of-dataset images



MR vs Pixel-RR



MR vs Pixel-RR



Comparison of different FES models sizes

Parameters	$\times 10^4$
StarGAN[2]	850
GANimation[3]	850
MR[4]	16.2
Pixel-KR	655
Pixel-RR	**3.28**

[2] Yunjey Choi et al. (2018). "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation." In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8789–8797.

[3] Albert Pumarola et al. (2020). "GANimation: One-shot anatomically consistent facial animation." In: *International Journal of Computer Vision* 128.3, pp. 698–713.

[4] Nazar Khan et al. (2020). "Masked Linear Regression for Learning Local Receptive Fields for Facial Expression Synthesis." In: *International Journal of Computer Vision* 128.5, pp. 1433–1454.

User study to evaluate expressions

Evaluators were asked to choose the best synthesized happy image considering

- perceptual quality
- expression realism
- identity preservation

Model	**Neutral $\rightarrow$ Happy**
GANimation	26%
MR	17%
Pixel-RR	**57%**

Expression classification accuracy

- A pre-trained[5] expression classifier is used to classify the synthesized happy images.

Model	Accuracy
GANimation	68%
MR	84%
Pixel-RR	**85%**

[5] https://github.com/thoughtworksarts/EmoPy

Conclusion

- We have presented a novel and simple Pixel-RR model for FES.

Conclusion

- We have presented a novel and simple Pixel-RR model for FES.
- Considers only one input pixel to produce an output pixel.

Conclusion

- We have presented a novel and simple Pixel-RR model for FES.
- Considers only one input pixel to produce an output pixel.
- Promising results on in-dataset and out-of-dataset images as compared to state-of-the-art GAN models.

Conclusion

- We have presented a novel and simple Pixel-RR model for FES.
- Considers only one input pixel to produce an output pixel.
- Promising results on in-dataset and out-of-dataset images as compared to state-of-the-art GAN models.
- Requires two orders of magnitude fewer parameters.

Conclusion

- We have presented a novel and simple Pixel-RR model for FES.
- Considers only one input pixel to produce an output pixel.
- Promising results on in-dataset and out-of-dataset images as compared to state-of-the-art GAN models.
- Requires two orders of magnitude fewer parameters.
- Can be deployed in mobile devices and embedded systems.

Conclusion

- We have presented a novel and simple Pixel-RR model for FES.
- Considers only one input pixel to produce an output pixel.
- Promising results on in-dataset and out-of-dataset images as compared to state-of-the-art GAN models.
- Requires two orders of magnitude fewer parameters.
- Can be deployed in mobile devices and embedded systems.

Thank you for your attention.

References I

Choi, Yunjey et al. (2018). "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation." In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8789–8797.

Khan, Nazar et al. (2020). "Masked Linear Regression for Learning Local Receptive Fields for Facial Expression Synthesis." In: *International Journal of Computer Vision* 128.5, pp. 1433–1454.

Pumarola, Albert et al. (2020). "GANimation: One-shot anatomically consistent facial animation." In: *International Journal of Computer Vision* 128.3, pp. 698–713.