SIGGRAPH ASIA 2018 TOKYO
CONFERENCE 4 – 7 December 2018
EXHIBITION 5 – 7 December 2018
Tokyo International Forum, Japan
SA2018.SIGGRAPH.ORG

Sponsored by

Deep Unsupervised Pixelization

Chu Han[1, 2, 3], Qiang Wen[2], Shengfeng He[2], Qianshu Zhu[2], Yinjie Tan[2], Guoqiang Han[2], Tien-Tsin Wong[1, 3]

[1]The Chinese University of Hong Kong,

[2]South China University of Technology,

[3]Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology, SIAT.

PIXEL ARTS

- Early gaming devices and computer system

©Nintendo

Final Fantasy ©SQUARE ENIX

Limited resolution and colors

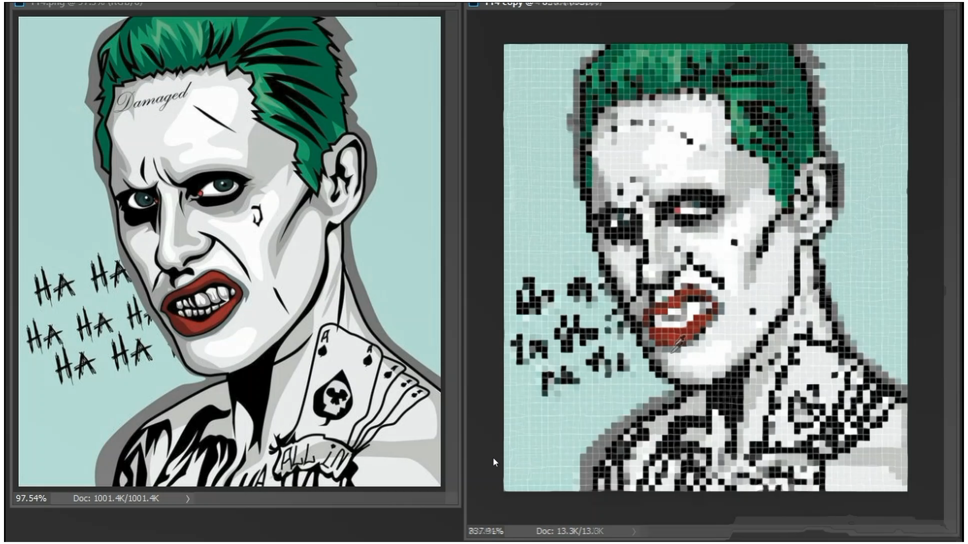
PIXEL ARTS

- Become an art form
  - Pixel art game
  - Portrait

CHALLENGING

- Manually designed pixel arts
  - Pixel-by-pixel
  - Tedious
  - Time consuming

x4 speed



RELATED WORKS

- Image Downscaling
  - Perceptually based [Öztireli and Gross 2015]
  - Detail-preserving [Weber et al. 2016]
  - Content adaptive [Kopf et al. 2013]

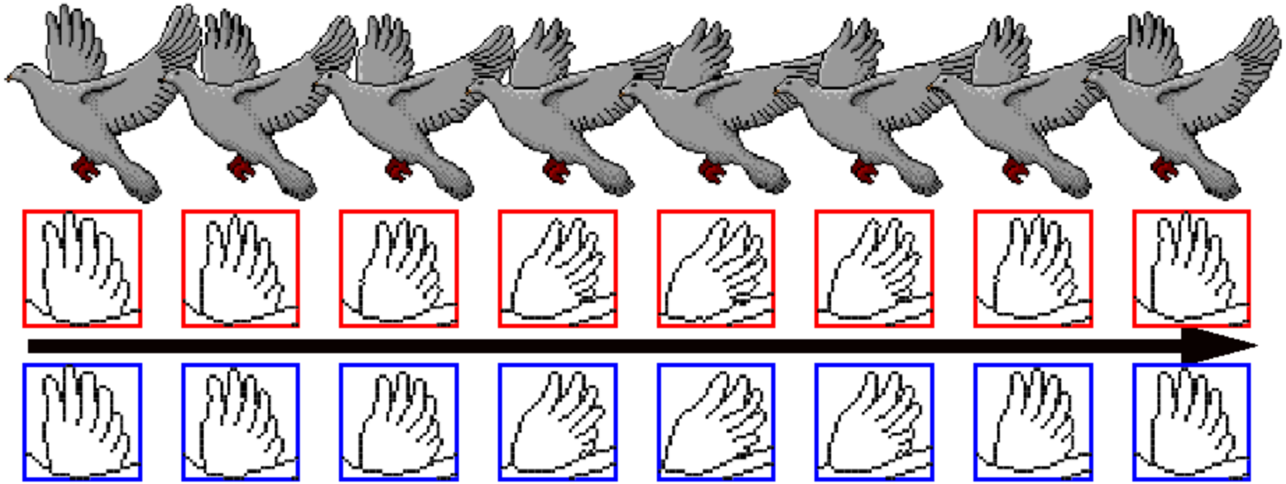
RELATED WORKS

- Image Downscaling
- Kernel-based nature can hardly synthesize sharp edges.

Input Perceptual (1/8) Content adaptive (1/8)

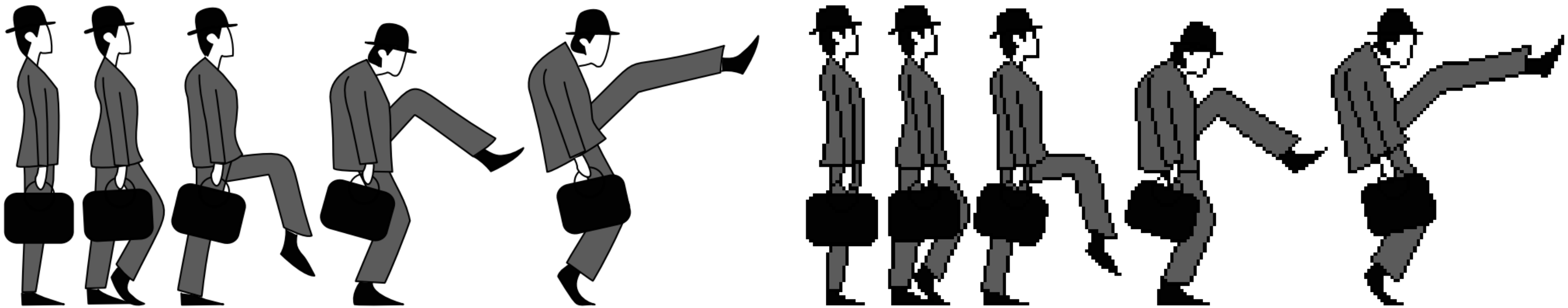
RELATED WORKS

- Optimization Approaches
  - Pixel art animation [Kuo et al. 2016]



RELATED WORKS

- Optimization Approaches
  - Pixel art animation [Kuo et al. 2016]
  - Rasterize vector line arts [Inglis et al. 2013]



(a) *Vector input* **(b)** *Pixelated output*

RELATED WORKS

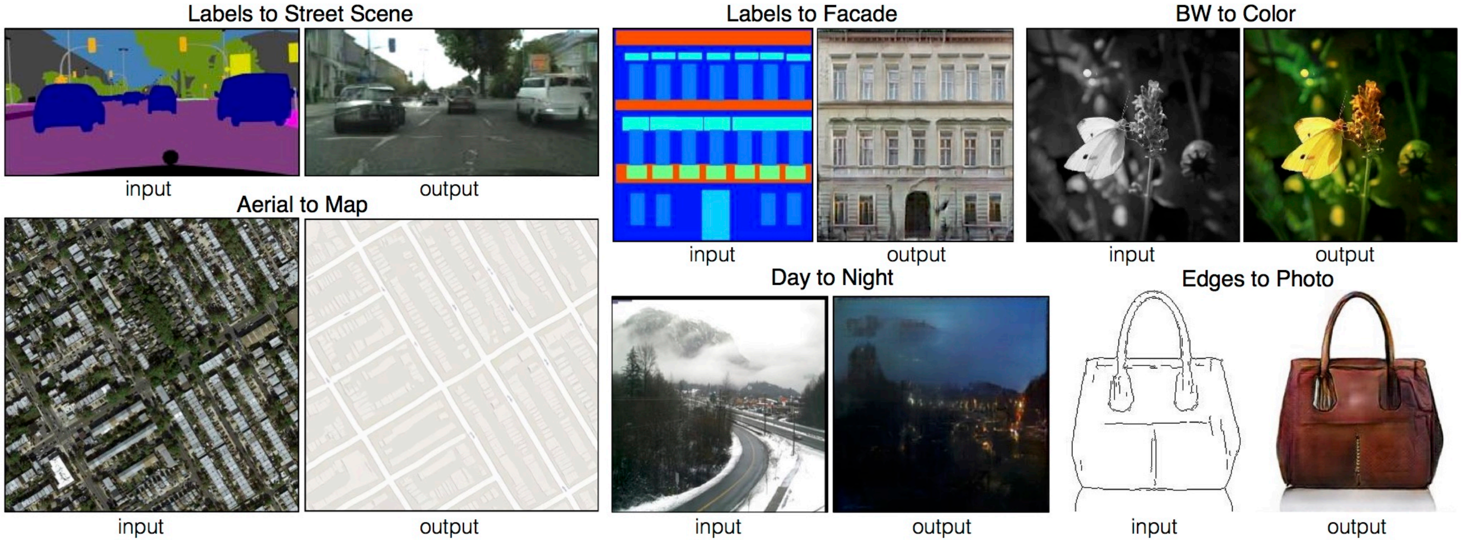
- Optimization Approaches
  - Pixel art animation [Kuo et al. 2016]
  - Rasterize vector line arts [Inglis et al. 2013]
  - Image abstraction [Gerstner et al. 2012]

RELATED WORKS

- Optimization Approaches
  - Pixel art animation [Kuo et al. 2016]
  - Rasterize vector line arts [Inglis et al. 2013]
  - Image abstraction [Gerstner et al. 2012]
- Pay more attention to the accuracy than the aesthetic consideration

RELATED WORKS

- Image-to-image translation
  - Labels to street scene [Isola et al. 2017]

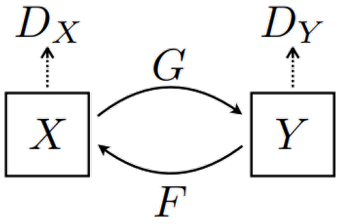


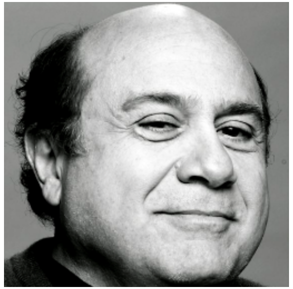
RELATED WORKS

- Image-to-image translation
  - Labels to street scene [Isola et al. 2017]
  - [Mirza and Osindero 2014]
  - [Odena et al. 2016]
  - [Xie and Tu 2015]
- Hard to collect paired training data of pixel arts

UNSUPERVISED LEARNING

- Cycle consistency loss
  - [Zhu et al. 2017]
  - [Yi et al. 2017]
- Artifacts and inconsistent colors

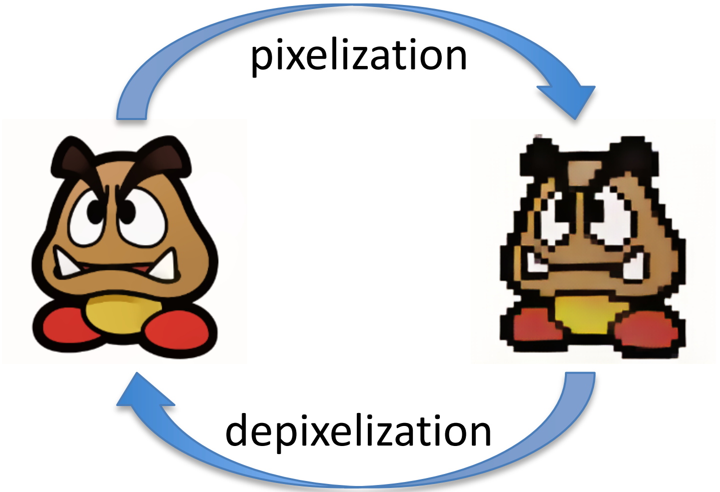



Input


CycleGAN

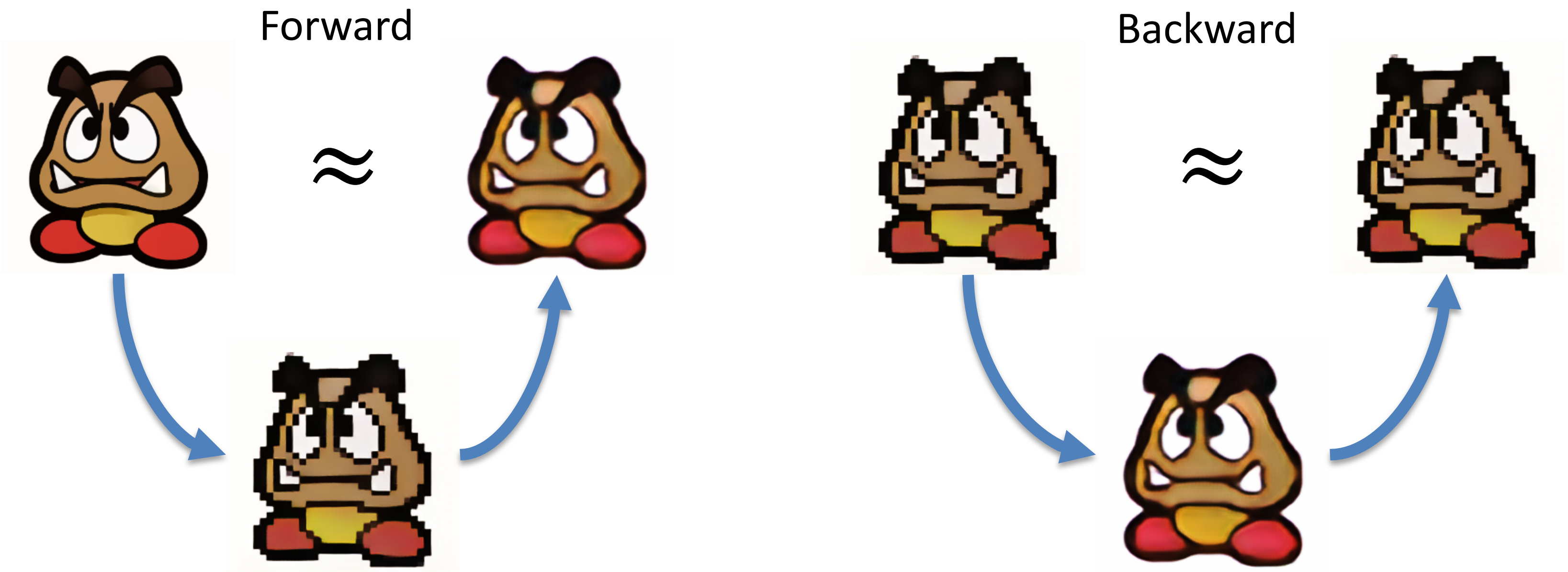
KEY IDEA

- Duality of pixelization and depixelization



OUR UNSUPERVISED LEARNING DESIGN

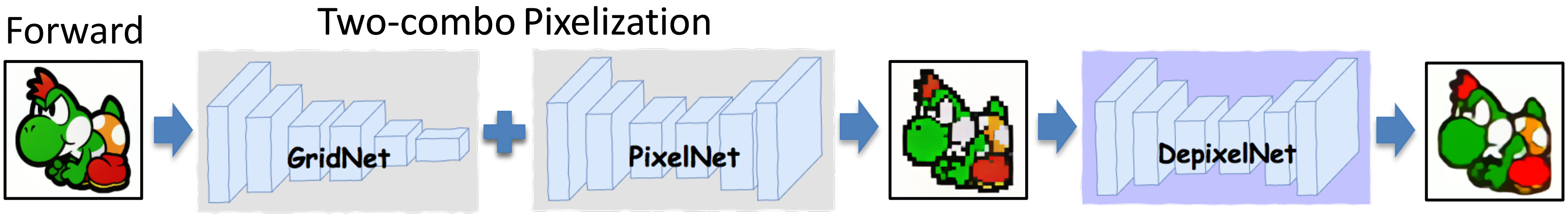
- Reversable training loop for unsupervised learning



Our approach

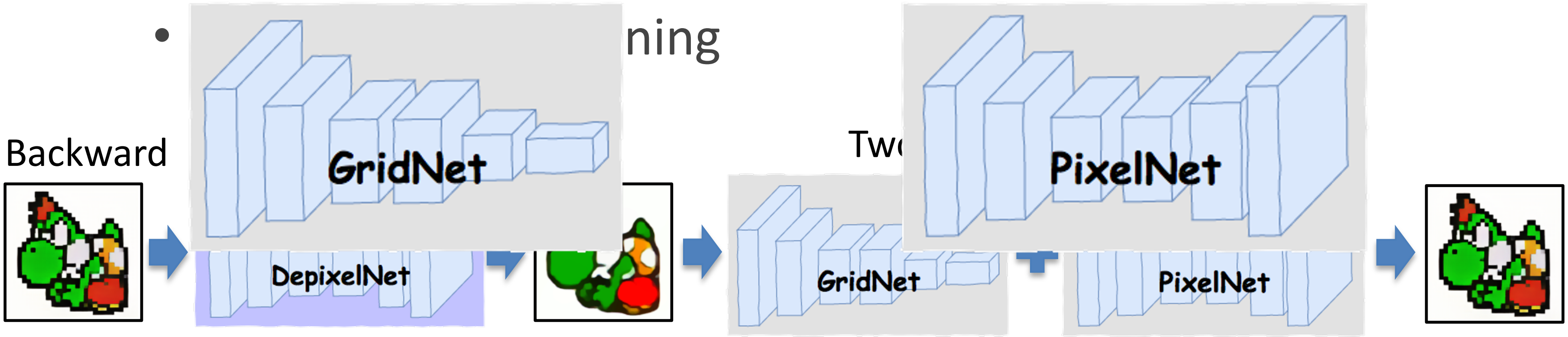
NETWORK STRUCTURE

- Cascaded network
  - Three subnetworks
  - Bi-directional training



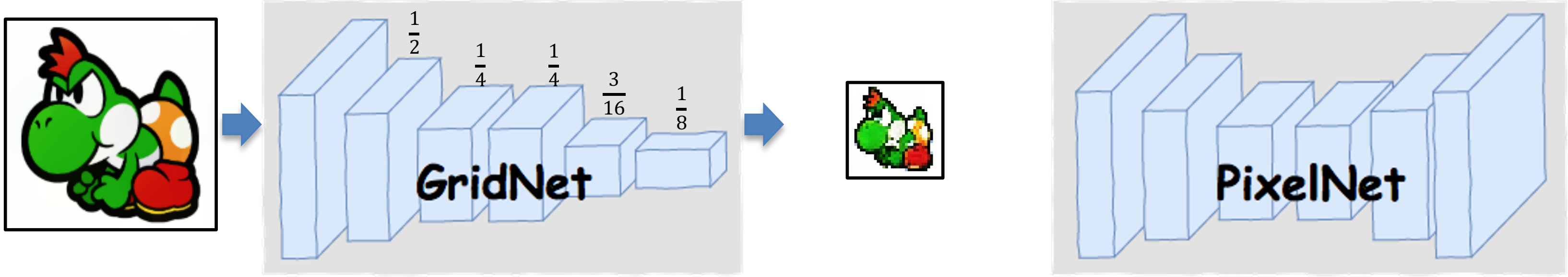
NETWORK STRUCTURE

- Cascaded network
  - Three subnetworks
  - ning



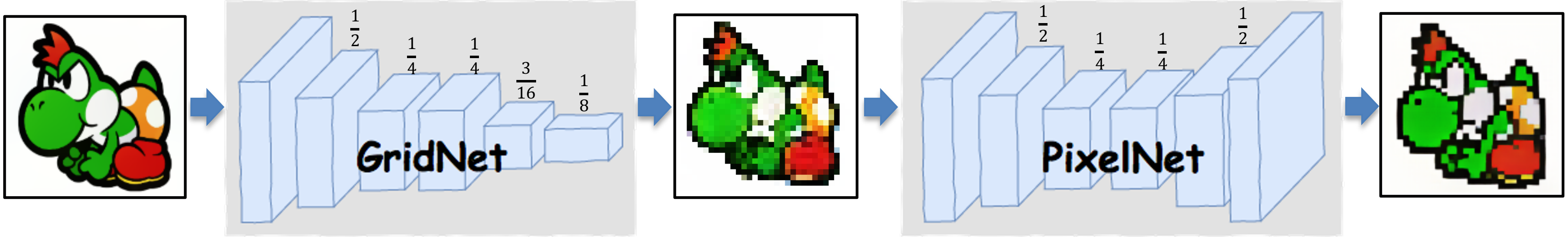
TWO-COMBO PIXELIZATION

- GridNet: initialize aliasing effect



TWO-COMBO PIXELIZATION

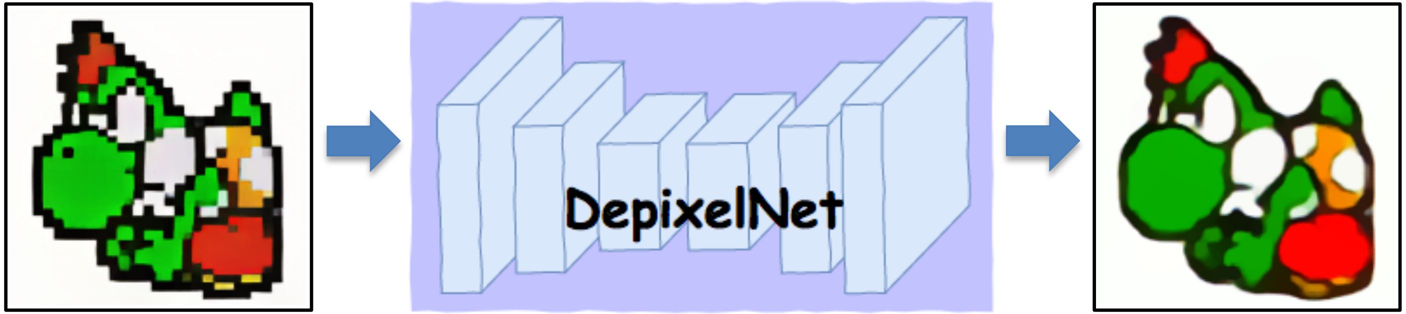
- PixelNet: refine pixel art and generate crisper edges



Easier for training and get a better result

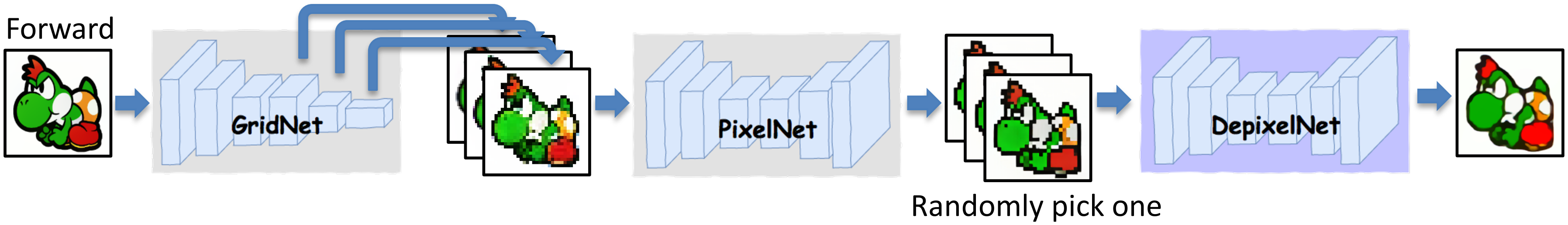
DEPIXELNET

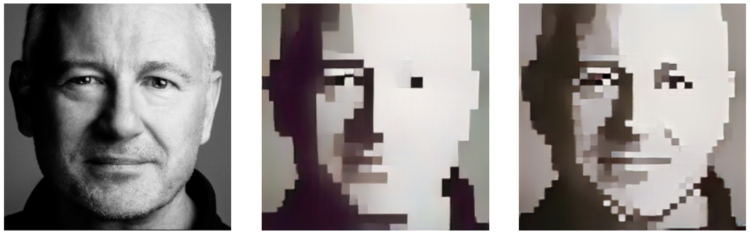
- Depixelize pixel arts



MULTISCALE TRAINING

- Improve the generalization





Input Single level Multi-level

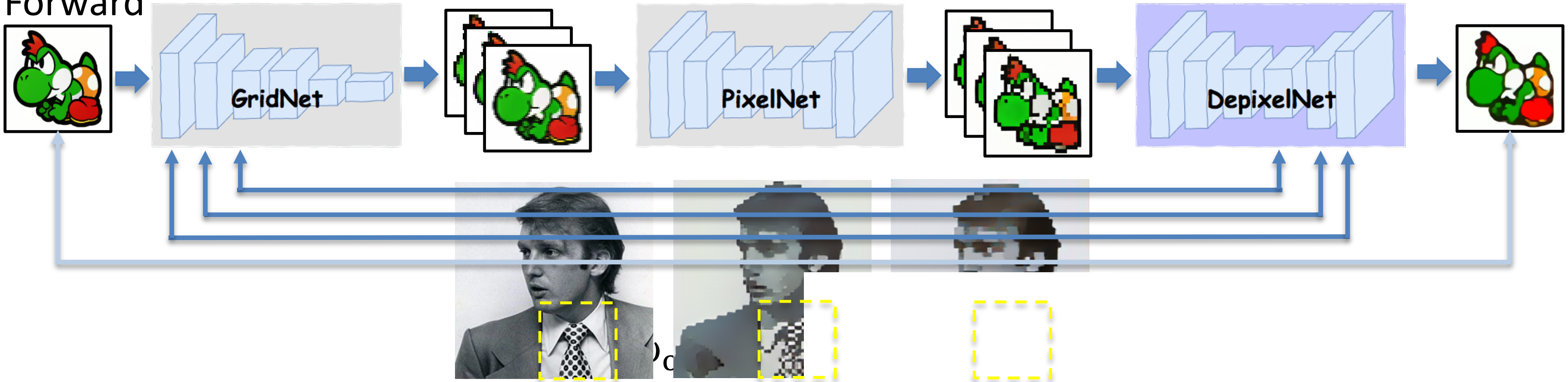
Allows network to learn cross-level semantically important details

LOSSES

- Mirror loss
- Adversarial loss
- L1 loss
- Gradient loss

MIRROR LOSS

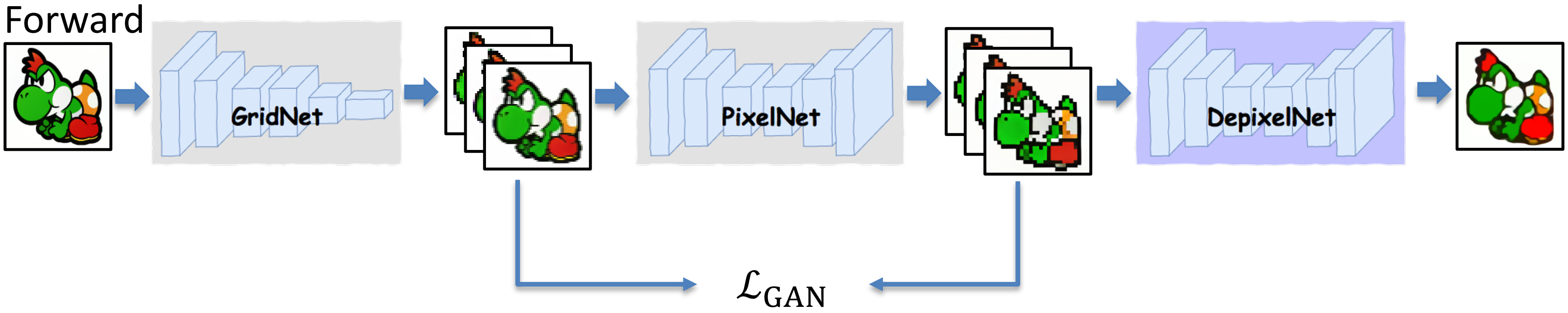
- Hold the reversibility of unsupervised learning
  - Input/output, f



Input w/o Mirror w/ Mirror

ADVERSARIAL LOSSES

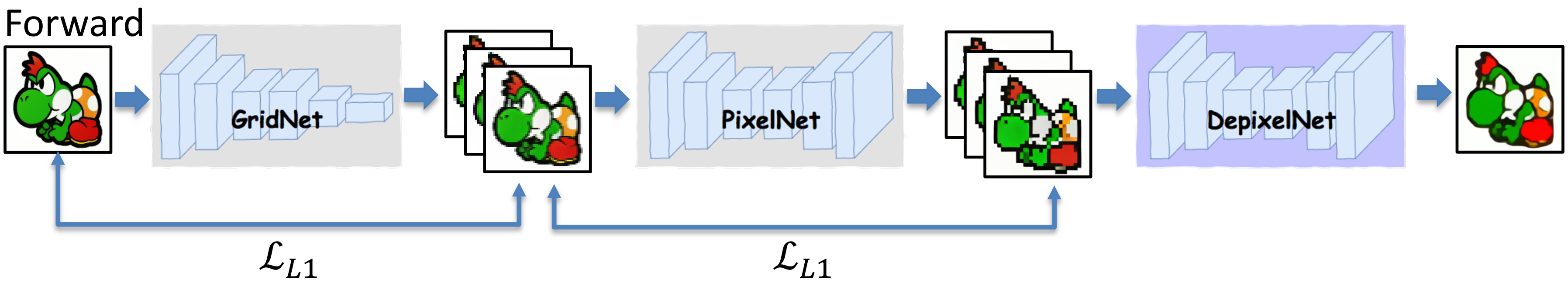
- Maintain pixel art style



Adversarial loss alone cannot guarantee the color correctness

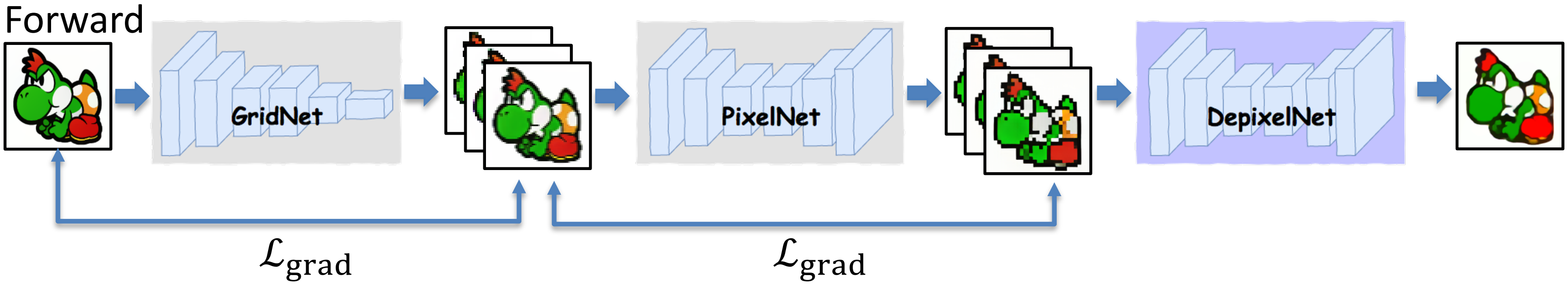
L_1 LOSSES

- Guarantee color consistency



GRADIENT LOSSES

- Ensure image smoothness and sharpness of edges



OBJECTIVE FUNCTION

- GridNet

$$\mathcal{L}_{\mathrm{GN}} = \mathcal{L}_{\mathrm{GAN}}(GN, \mathcal{D}_{\mathrm{GN}}, F) + \mathcal{L}_{L1\&\mathrm{grad}}(GN, F) + \mathcal{L}_{L1\&\mathrm{grad}}(GN, B)$$

- PixelNet

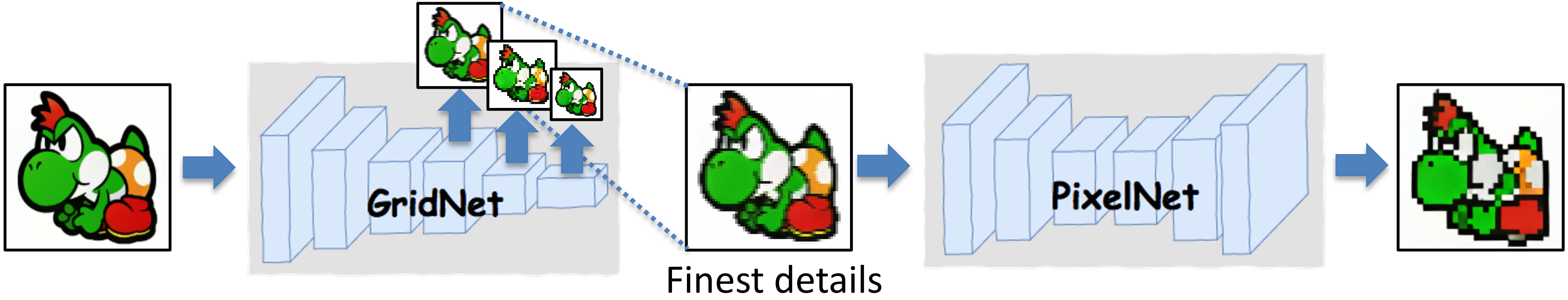
$$\mathcal{L}_{\mathrm{PN}} = \mathcal{L}_{\mathrm{GAN}}(PN, \mathcal{D}_{\mathrm{PN}}, F) + \mathcal{L}_{L1\&\mathrm{grad}}(PN, F) + \mathcal{L}_{\mathrm{mirr}}(DN \to PN, B)$$

- DepixelNet

$$\mathcal{L}_{\mathrm{DN}} = \mathcal{L}_{\mathrm{GAN}}(DN, \mathcal{D}_{\mathrm{DN}}, B) + \mathcal{L}_{L1\&\mathrm{grad}}(DN, B) + \mathcal{L}_{\mathrm{mirr}}(GN \to DN, F)$$

TESTING PHASE

- Training: three scales
- Testing: only output the third last conv-block
- Appearance: approximately 1/6 original input



Sponsored by

TRAINING DATA

- 900 pixel arts and 900 cliparts

Pixel arts

Cliparts

Results and experiments

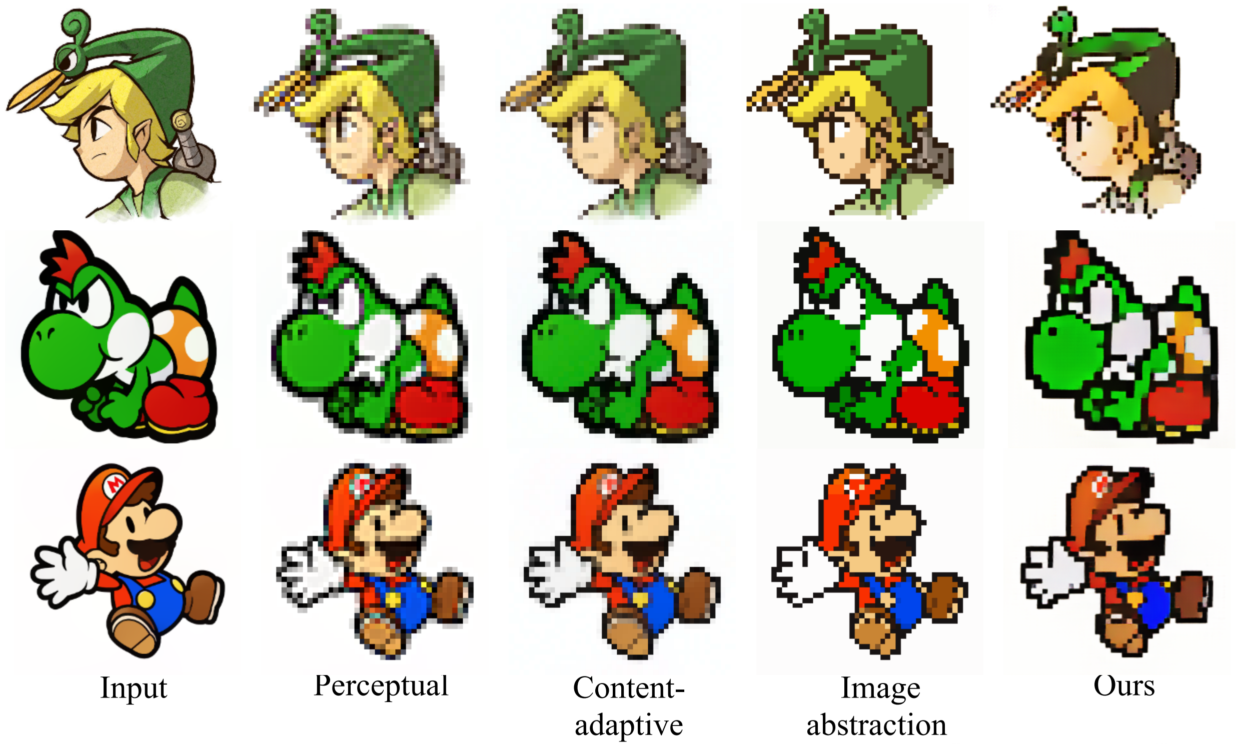
COMPETITORS

- Bicubic
- Perceptual [Öztireli and Gross 2015]
- Content-adaptive [Kopf et al. 2013]
- Image abstraction [Gerstner et al. 2012]

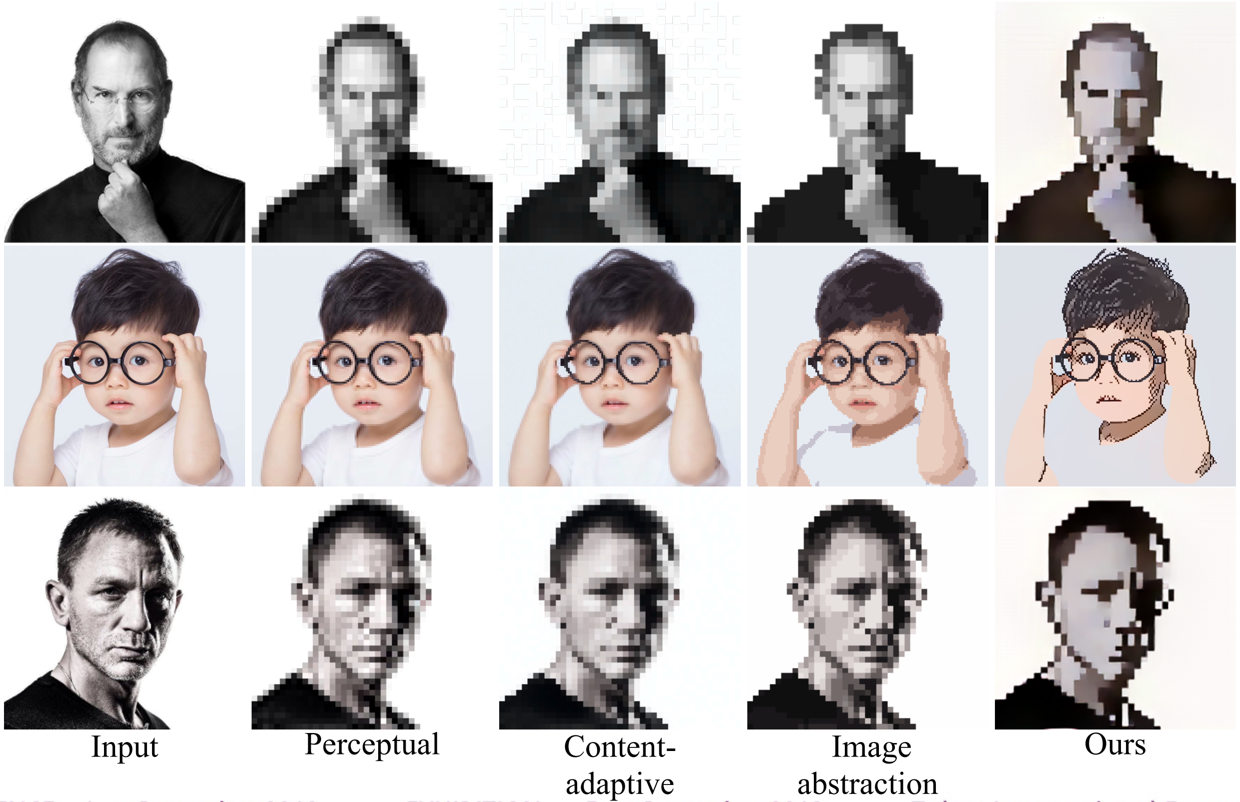
COMPARISONS TO EXISTING METHODS

Input Bicubic Perceptual Content-adaptive Ours

MORE RESULTS - CARTOON

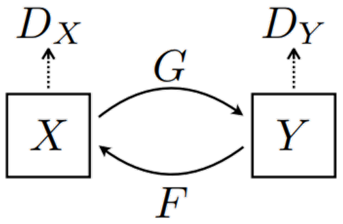


MORE RESULTS - PORTRAIT

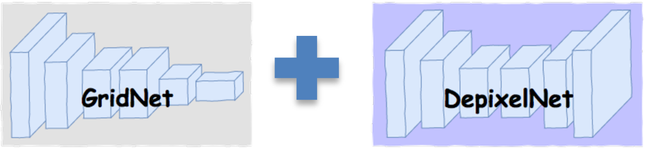


COMPARISON TO ALTERNATIVE CNN MODELS

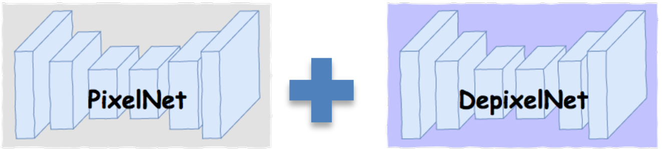
- CycleGan



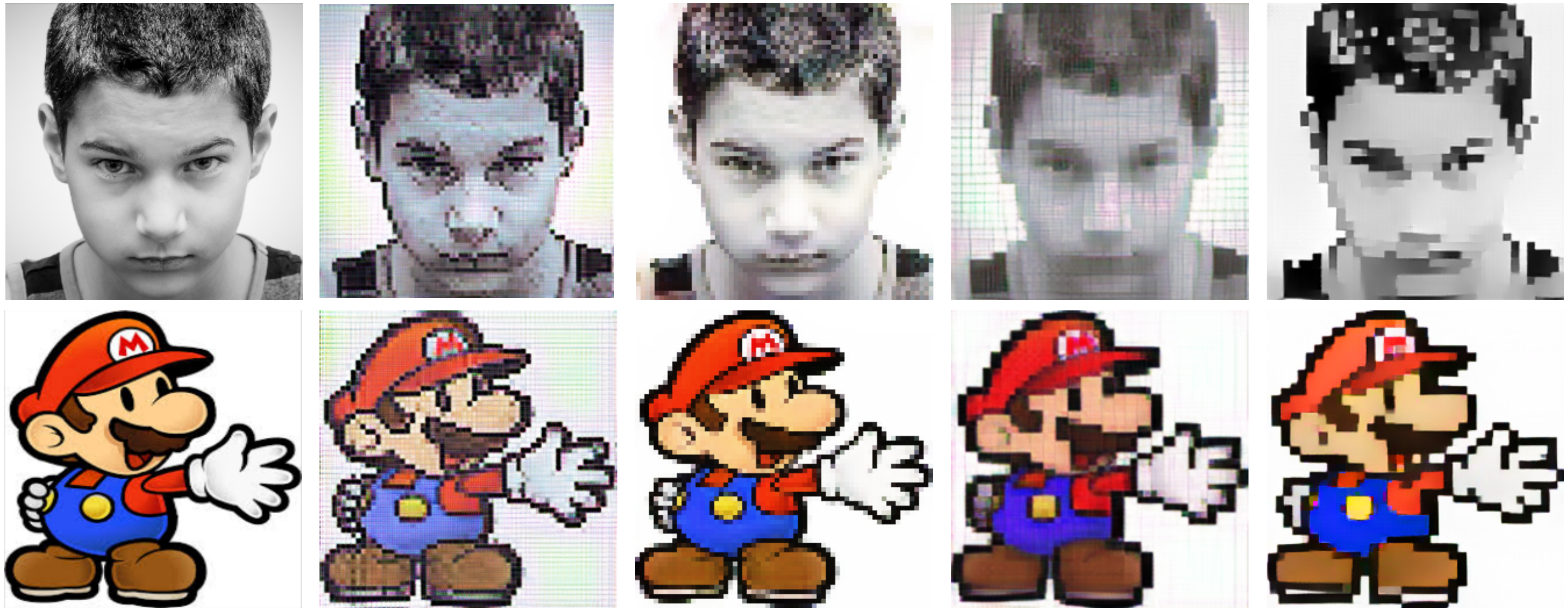
- “GridNet alone”



- “PixelNet alone”



COMPARISON TO ALTERNATIVE CNN MODELS

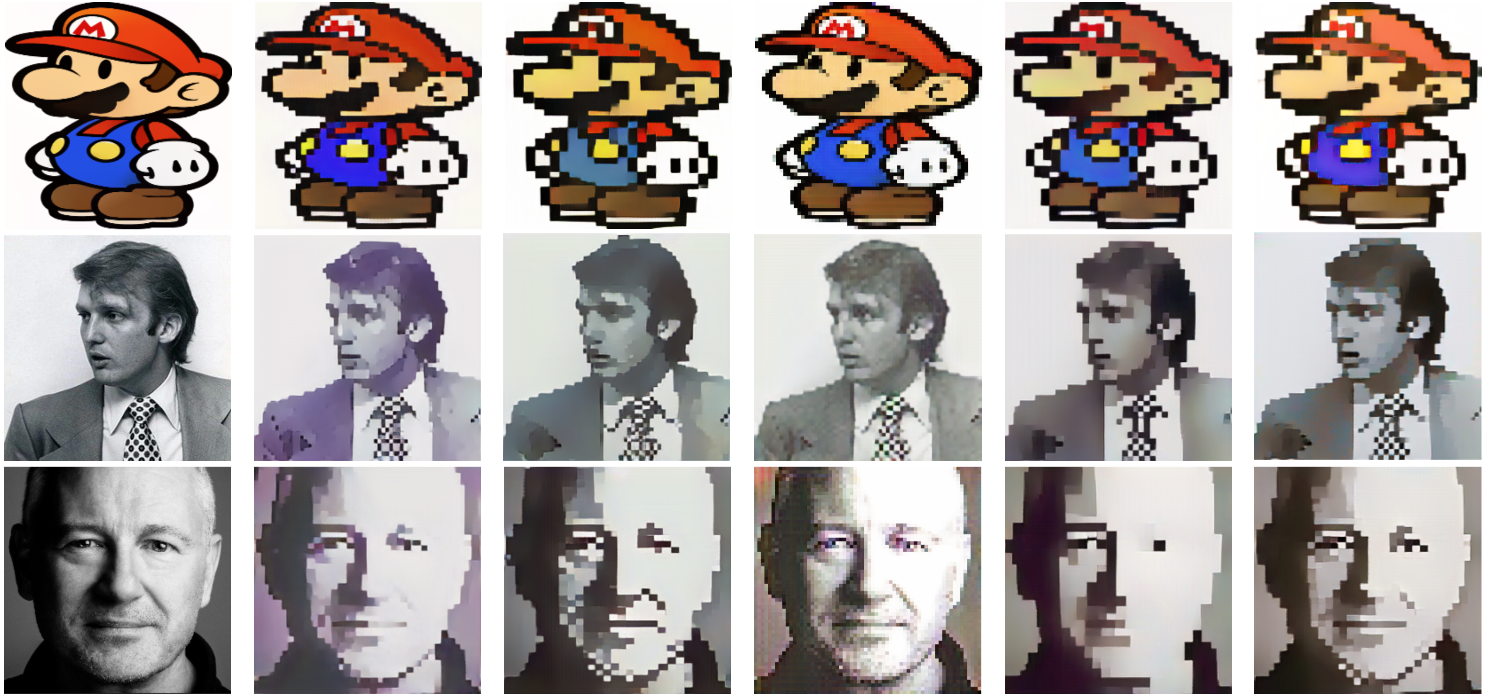


Input | CycleGAN | “GridNet alone” | “PixelNet alone” | Complete model

IMPACT OF LOSSES

- Loss1: $L_{L1} + L_{\text{mirr}} + L_{\text{GAN}}$
  Loss2: $L_{L1} + L_{\text{grad}} + L_{\text{GAN}}$
  Loss3: $L_{L1} + L_{\text{grad}} + L_{\text{mirr}}$
  Loss4: $L_{L1} + L_{\text{grad}} + L_{\text{mirr}} + L_{\text{GAN}}$ (all w/o multi-scale)
  Loss5: $L_{L1} + L_{\text{grad}} + L_{\text{mirr}} + L_{\text{GAN}}$ (all w/ multi-scale)

IMPACT OF LOSSES



Input	Loss1 (w/o Gradient)	Loss2 (w/o Mirror)	Loss3 (w/o GAN)	Loss4 (w/o Multi-scale)	Loss5 (All)

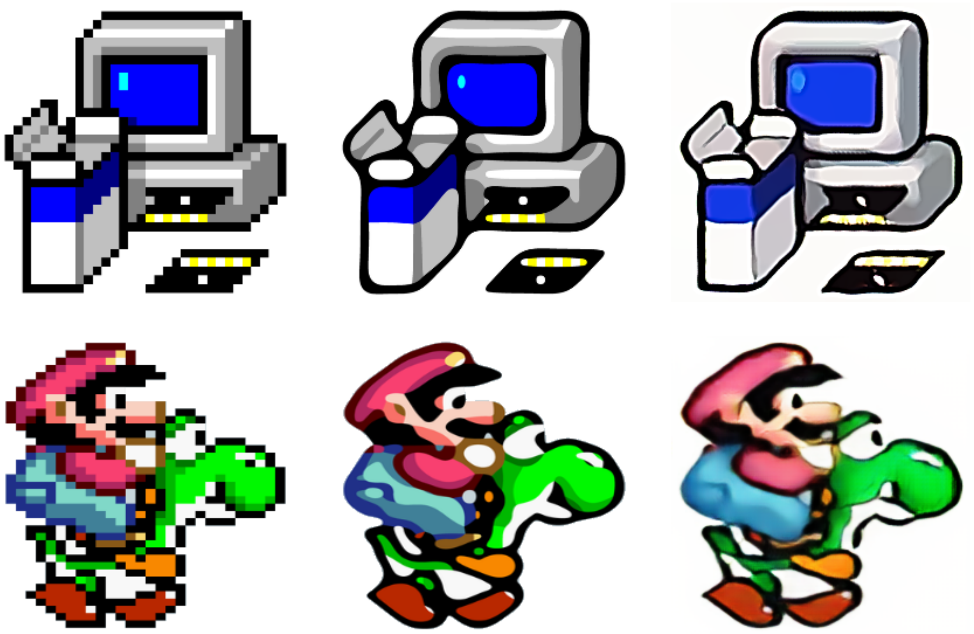
COMPARISON TO MANUAL PIXEL ARTS

Input

Manual pixel art
©Vixels

Ours (Network output)

DEPIXELIZATION



Input [Kopf and Lischinski 2011] Ours

LIMITATIONS

- Pixelized appearance is always approximately 1/6 of the resolution of the input
- Unpredictable artifacts and color change introduced by GAN

Input

Ours

CONCLUSIONS

- In this paper, we propose a cascaded network for unsupervised pixelization.
- Mirror loss is proposed to hold the reversibility of our unsupervised learning.
- Dividing the network into three subnetworks is more effective than solving with a generic network.

Sponsored by

SIGGRAPH ASIA 2018 TOKYO

CONFERENCE 4 – 7 December 2018
EXHIBITION 5 – 7 December 2018
Tokyo International Forum, Japan
SA2018.SIGGRAPH.ORG

THANK YOU!

The Chinese University of Hong Kong

Email: chan@cse.cuhk.edu.hk

South China University of Technology

Email: shengfenghe7@gmail.com