Towards Object Shape Translation Through Unsupervised Generative Deep Models

Lies Bollens, Tinne Tuytelaars & José Oramas M.

KU Leuven, ESAT-PSI

September 25th, 2019

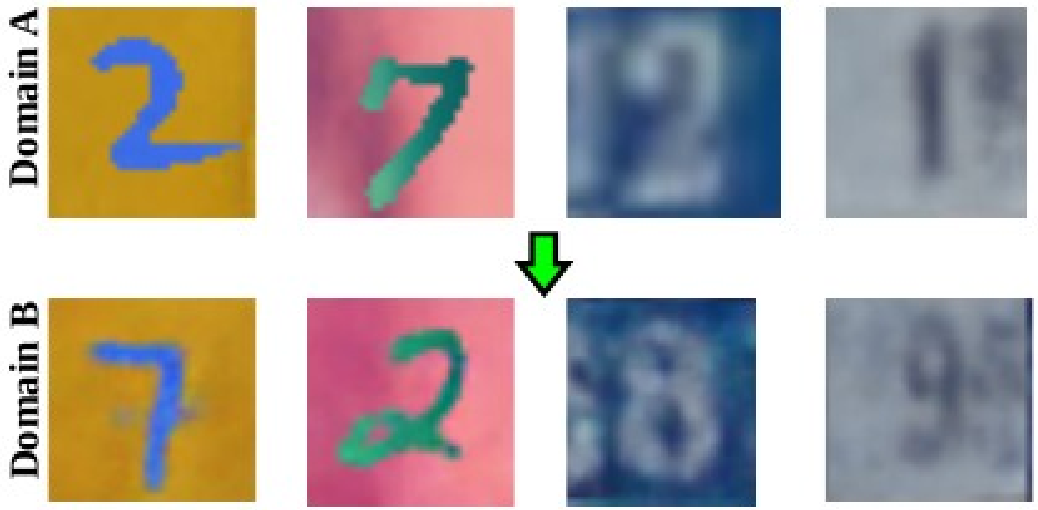
In a nutshell ...

What?

Translating shape, preserving style (colour, texture, etc)

input

output



e.g. across
digit classes

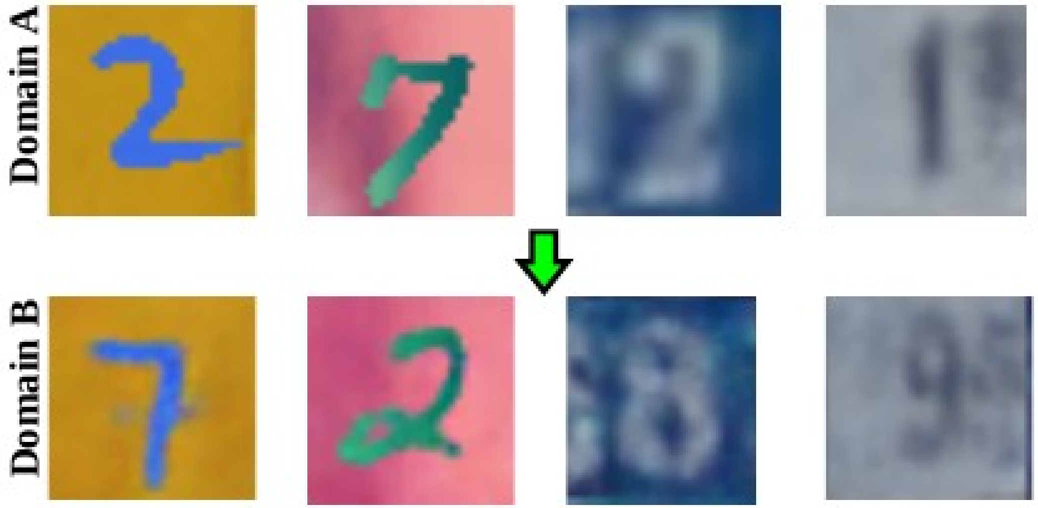
In a nutshell ...

What?

Translating shape, preserving style (colour, texture, etc)

input

output



e.g. across digit classes

How?

- Learn shapes independently
- Learn how to map the learned shapes

Some history

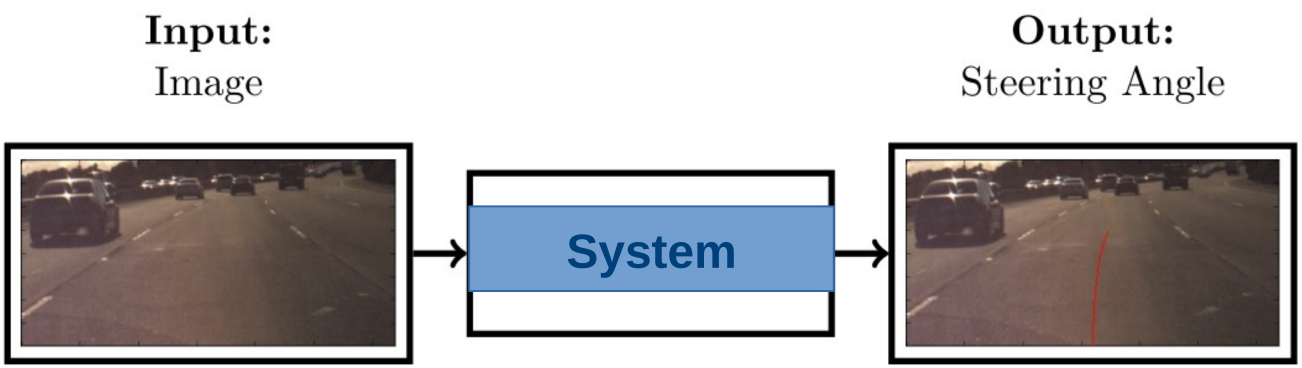
End-to-End
Steering Prediction

(2016 - 2017)

Previous Work ...

End-to-end Steering Prediction[1]

Given an image sequence → Predict a steering angle



[1]Heylen et al., "From Pixels to Actions: Learning to Drive a Car with Deep Neural Networks". WACV'18.

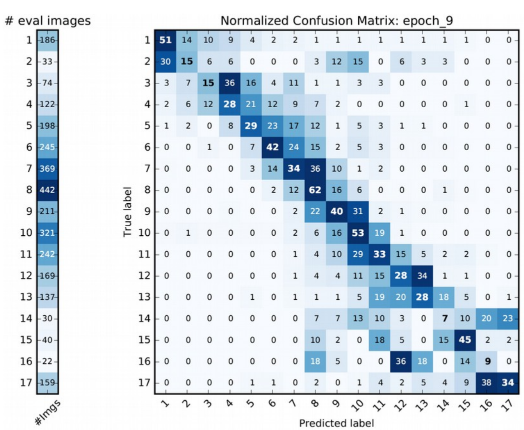
Previous Work ...

End-to-end Steering Prediction[1]

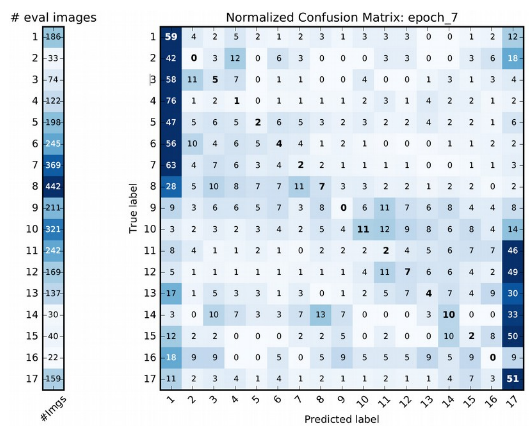
Some Results

Data: Udacity & GTA-V Simulator

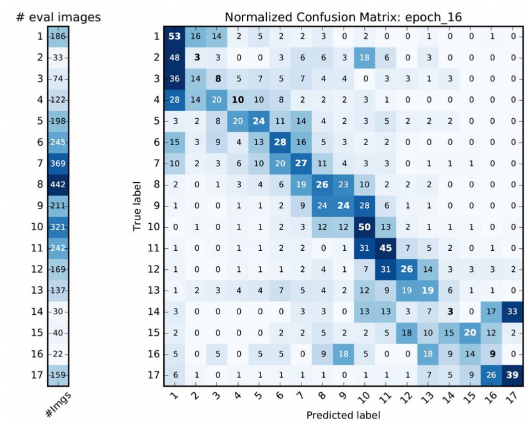
Regular only



Recovery only



Regular + Recovery



worst **best**

[1]Heylen et al., "From Pixels to Actions: Learning to Drive a Car with Deep Neural Networks". WACV'18.

KU LEUVEN

Nice, but ...

Training Conditions (simulator)

≠

Testing Conditions (KITTI dataset)

Far from the testing conditions

Making simulation data more realistic

(2017 - 2018)

From the virtual world to reality

What?

Given simulated data → Bring it closer to testing conditions

Grand Theft Auto V

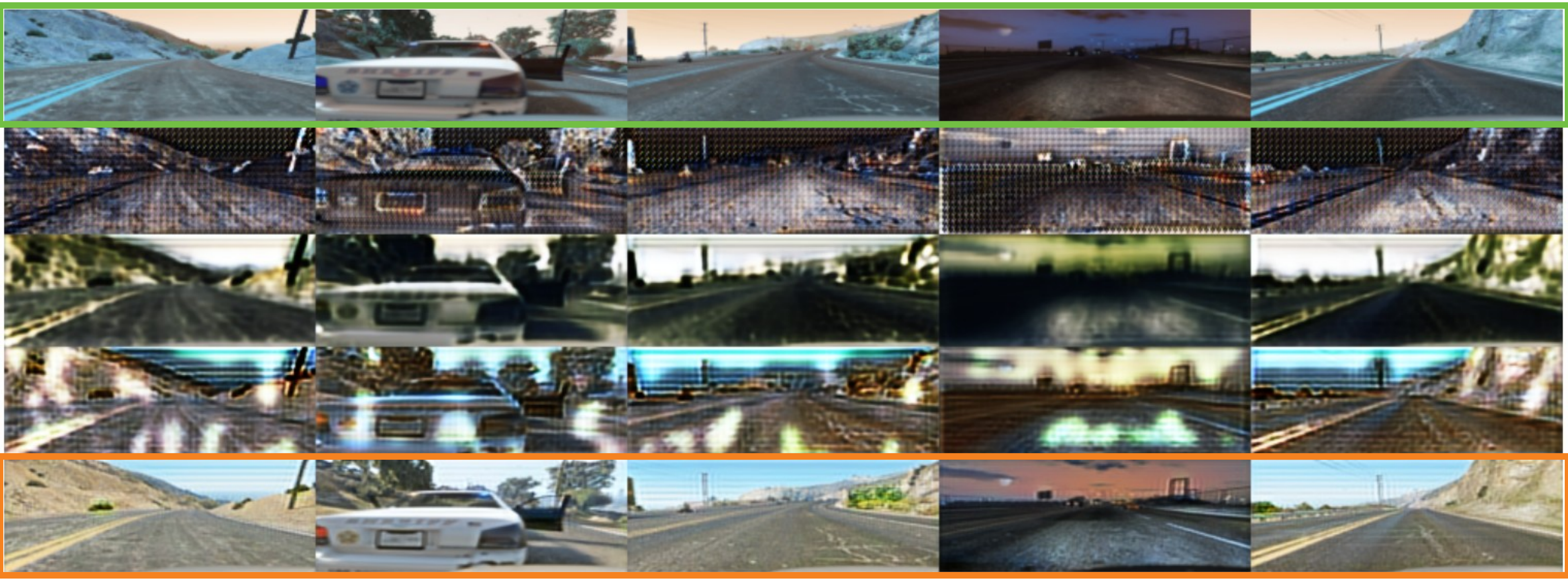
Example from the KITTI dataset

How?

- Domain/image translation
- Generative Adversarial Networks (GANs)[1]

Goodfellow et al., "Generative Adversarial Nets," in NIPS 2014.

From GTA-V to realistic images



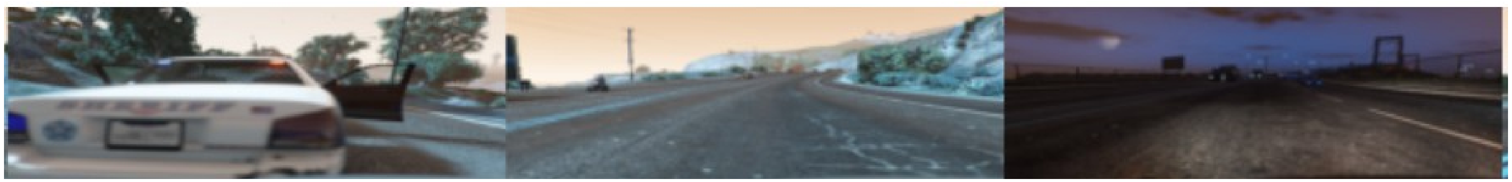
- Original image examples from GTA-V (*input*)
- GAN + Wasserstein Loss

KU LEUVEN

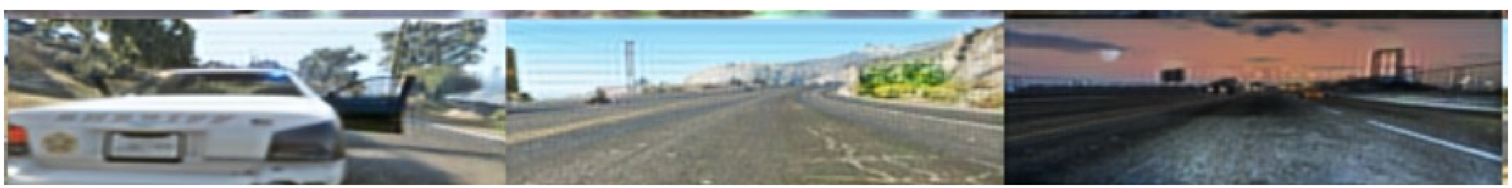
Nice but...

Currently it is mostly about changing pixel colours,
What about the structure?
(shape of trees, models of cars, building architecture)

input



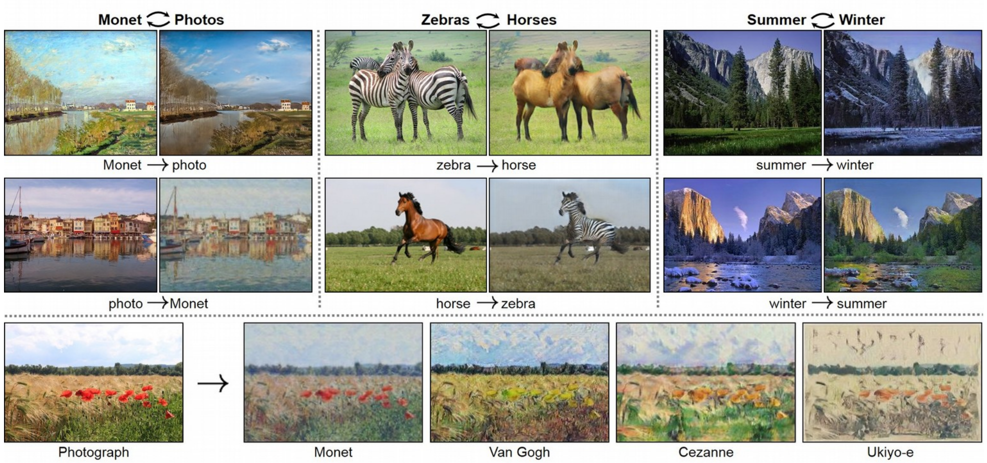
output



Translating Shapes, Preserving Style

Related Work

Neural Style Transfer

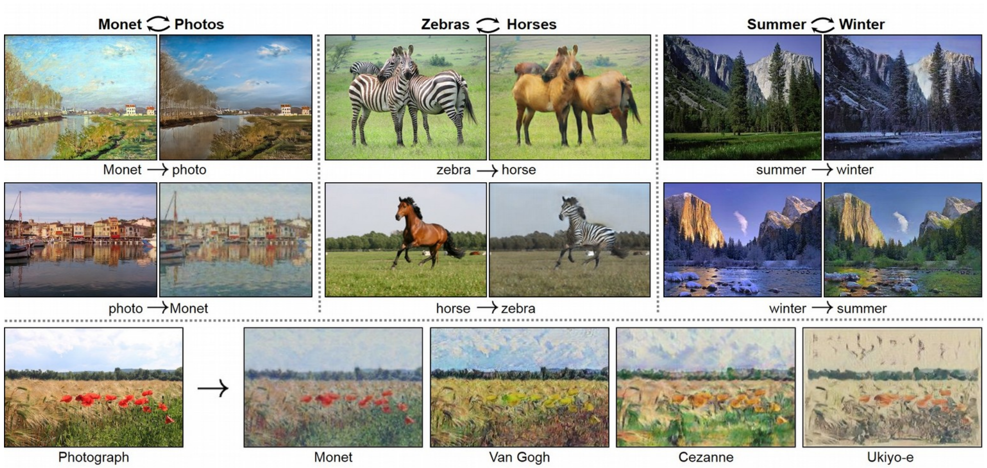


Zhu et al., ICCV'17.

- Gatys et al., CVPR'16.
- Zhu et al., ICCV'17.
- Luan et al., CVPR'17.
- Mechrez et al., BMVC'17.
- Wang et al., CVPR'17
- Liu et al., NPAR'17,
- Dumoulin et al., ICLR'17.
- Chen et al., CVPR'17.
- Li et al. CVPR'17.
- Zhang et al., arXiv:1703.06953.
- Jing et al., ECCV'18.
- **Jing et al., arXiv:1705.04058.**

Related Work

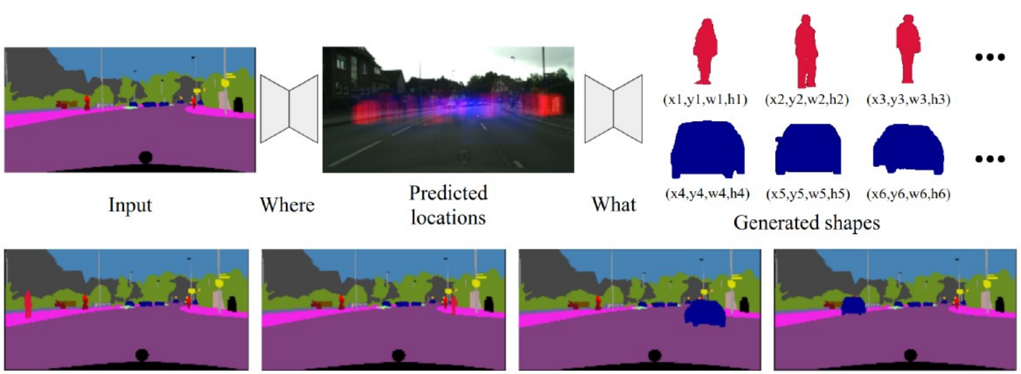
Neural Style Transfer



Zhu et al., ICCV'17.

- Gatys et al., CVPR'16.
- Zhu et al., ICCV'17.
- Luan et al., CVPR'17.
- Mechrez et al., BMVC'17.
- Wang et al., CVPR'17
- Liu et al., NPAR'17,
- Dumoulin et al., ICLR'17.
- Chen et al., CVPR'17.
- Li et al. CVPR'17.
- Zhang et al., arXiv:1703.06953.
- Jing et al., ECCV'18.
- **Jing et al., arXiv:1705.04058.**

Shape Transfer



Lin et al., CVPR'18.

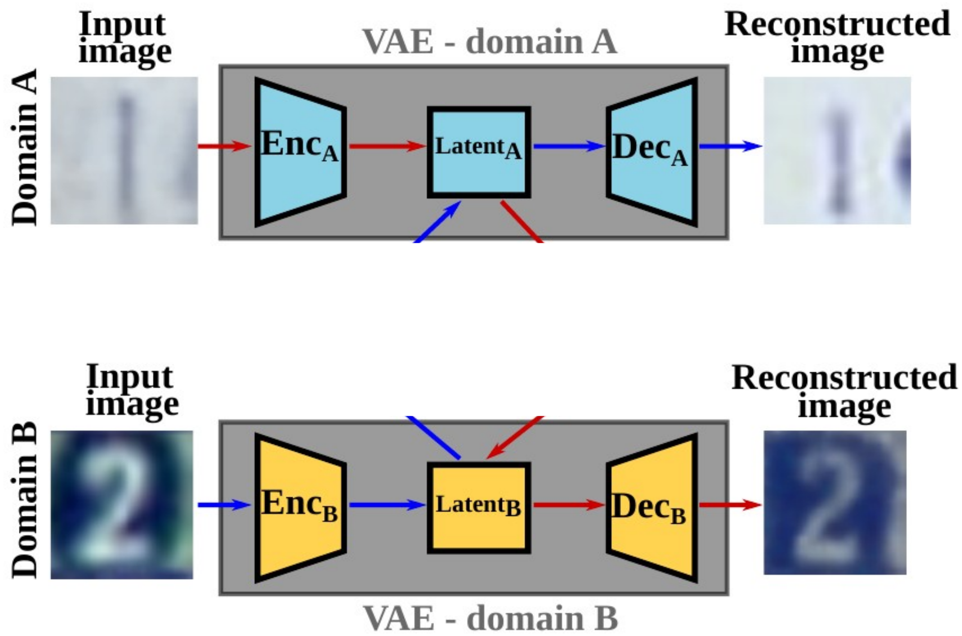
- Jaderberg et al., NIPS'15.
- Lee et al., NIPS'18.
- Lin et al., CVPR'18.

Proposed Method

Proposed Method

How? (notion)

1- Learn how to model each domain independently.

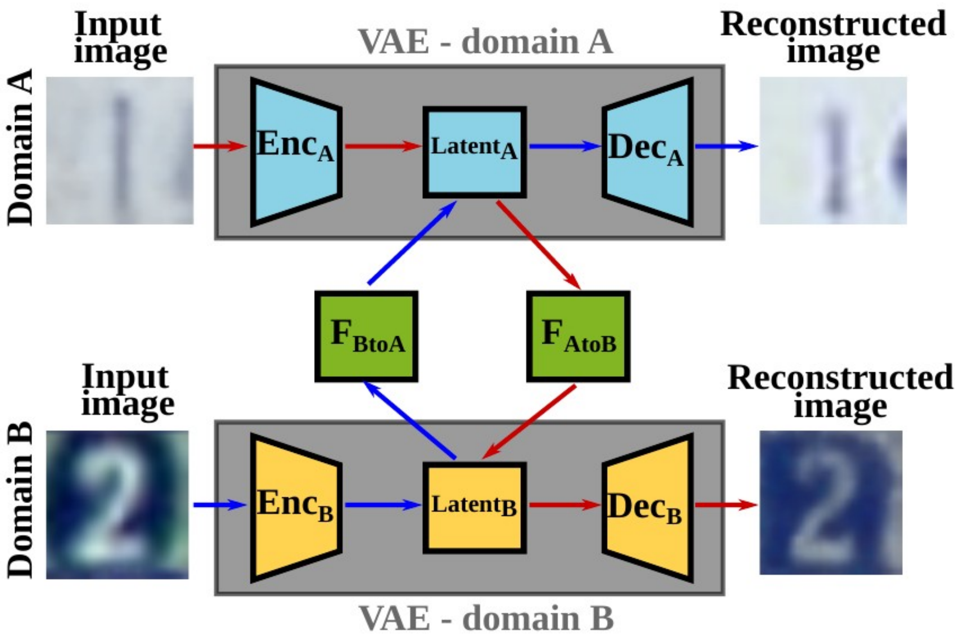


Proposed Method

How? (notion)

1- Learn how to model each domain independently.

2- Learn a translation function between the domains[1].



[1]Lejeune et al. , "A Data Driven Similarity Measure and Example Mapping Function for General, Unlabelled Data Sets", ECAI'16.

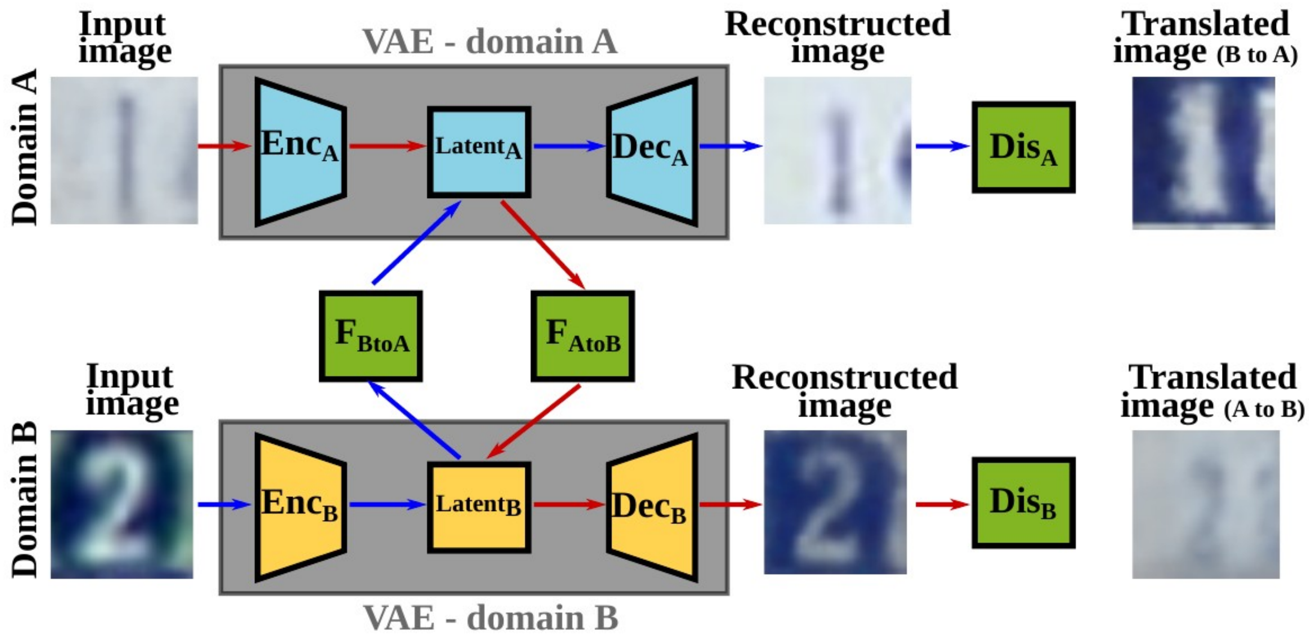
KU LEUVEN

Proposed Method

How? (notion)

1- Learn how to model each domain independently.

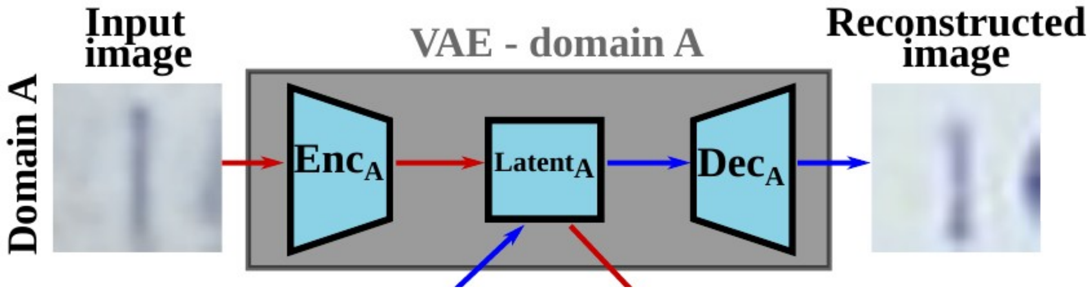
2- Learn a translation function between the domains.



Proposed Method

Modelling Domain Information

- Learn how to model each domain independently.
  - → Through variational autoencoders (VAE)[1]

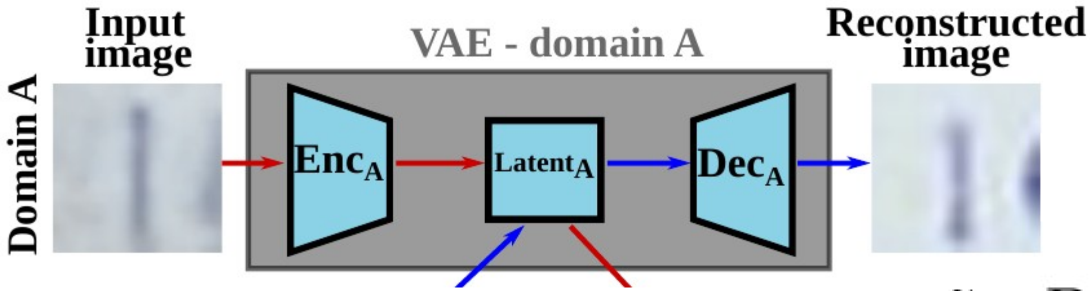


[1]D. P Kingma and M. Welling, “Auto-Encoding Variational Bayes,” in arXiv:1312.6114, 2013.

Proposed Method

Modelling Domain Information

- Learn how to model each domain independently.
  → Through variational autoencoders (VAE)[1]



x

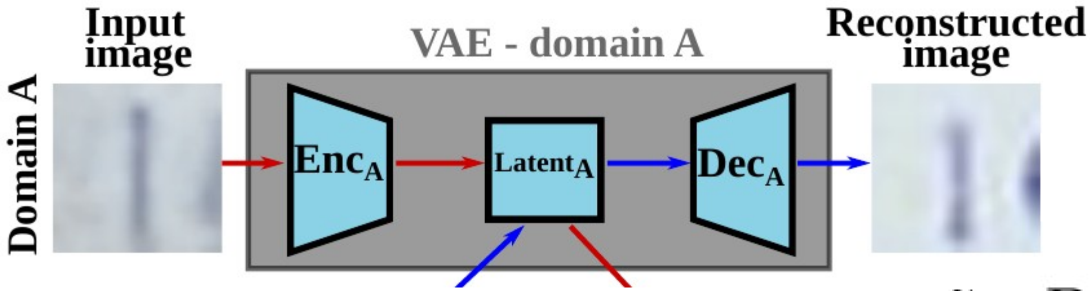
$$\tilde{x} \sim Dec(z) = p(x|z)$$

[1]D. P Kingma and M. Welling, "Auto-Encoding Variational Bayes," in arXiv:1312.6114, 2013.

Proposed Method

Modelling Domain Information

- Learn how to model each domain independently.
    → Through variational autoencoders (VAE)[1]



x

$$\tilde{x} \sim Dec(z) = p(x|z)$$

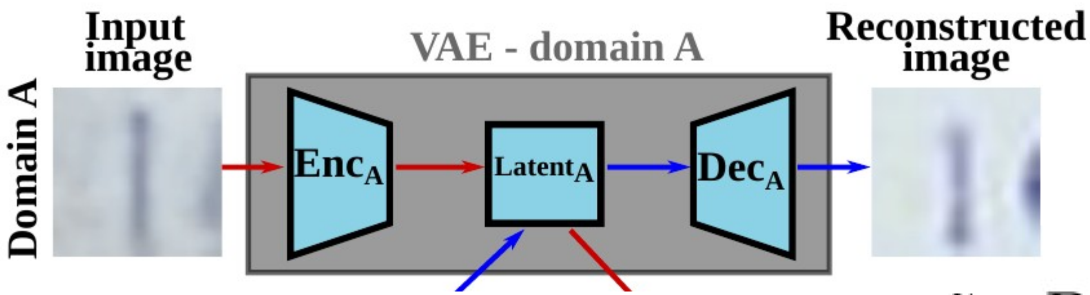
With latent space → $z \sim Enc(x) = q(z|x)$

[1]D. P Kingma and M. Welling, “Auto-Encoding Variational Bayes,” in arXiv:1312.6114, 2013.

Proposed Method

Modelling Domain Information

- Learn how to model each domain independently.
  → Through variational autoencoders (VAE)[1]



x

$$\tilde{x} \sim Dec(z) = p(x|z)$$

With latent space → $z \sim Enc(x) = q(z|x)$

Applying the Loss

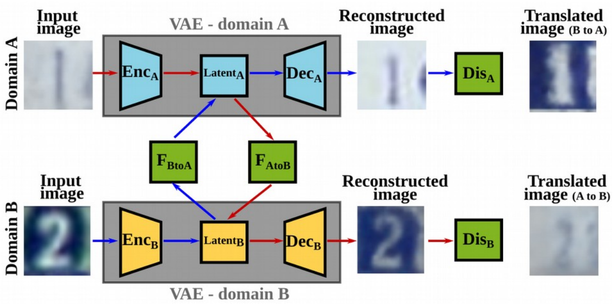
$$\mathcal{L}_{VAE} = -\mathbb{E}_{q(z|x)}\left[\log p(x|z)\right] + \mathcal{D}_{KL}\left(q(z|x)||p(z)\right)$$

[1]D. P Kingma and M. Welling, “Auto-Encoding Variational Bayes,” in arXiv:1312.6114, 2013.

Proposed Method

Shape Translation

- Learn a mapping function between the [shape] domains.
  → Through an extended CycleGAN[1.]



→ Applying the Loss

$$\mathcal{L}=\mathcal{L}_{GAN}(Gen_A, Dis_A)+\mathcal{L}_{GAN}(Gen_B, Dis_B)$$
$$+\lambda_{cycle} * \mathcal{L}_{cyc}(Gen_A, Gen_B)$$
$$+\lambda_{sim} * \mathcal{L}_{sim}(x, y, Gen_A, Gen_B)$$

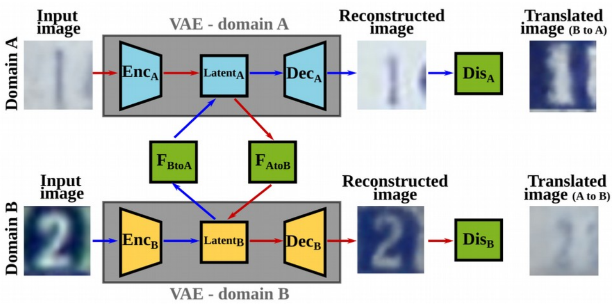
Zhu, et al. "Unpaired image- to-image translation using cycle-consistent adversarial networks", ICCV'17.

Proposed Method

Shape Translation

- Learn a mapping function between the [shape] domains.

→ Through an extended CycleGAN[1].



→ Applying the Loss

$$\mathcal{L} = \boxed{\mathcal{L}_{GAN}(Gen_A, Dis_A)} + \boxed{\mathcal{L}_{GAN}(Gen_B, Dis_B)} + \lambda_{cycle} * \mathcal{L}_{cyc}(Gen_A, Gen_B) + \lambda_{sim} * \mathcal{L}_{sim}(x, y, Gen_A, Gen_B)$$

Where:

$$\mathcal{L}_{GAN}(Gen_A, Dis_A) = \mathbb{E}_{x \in dom(A)}\left[log(Dis_A(x))\right] + \mathbb{E}_{y \in dom(B)}\left[log(1 - Dis_A(Gen_A(y)))\right]$$

→ ensures the mapping between latent spaces is accurate.

Zhu, et al. “Unpaired image- to-image translation using cycle-consistent adversarial networks”, ICCV’17.

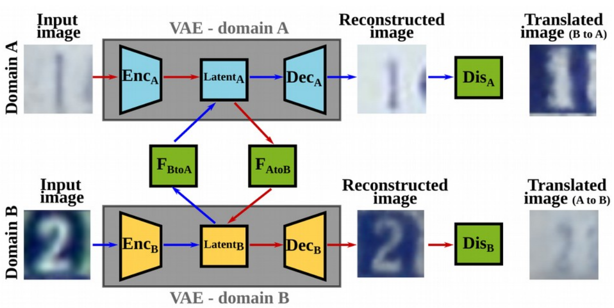
KU LEUVEN

Proposed Method

Shape Translation

- Learn a mapping function between the [shape] domains.

→ Through an extended CycleGAN[1.]



→ Applying the Loss

$$\mathcal{L} = \mathcal{L}_{GAN}(Gen_A, Dis_A) + \mathcal{L}_{GAN}(Gen_B, Dis_B) + \lambda_{cycle} * \boxed{\mathcal{L}_{cyc}(Gen_A, Gen_B)} + \lambda_{sim} * \mathcal{L}_{sim}(x, y, Gen_A, Gen_B)$$

Where:

$$\mathcal{L}_{cyc} = \mathbb{E}_{x \in dom(A)} \left[||Gen_B(Gen_A(x)) - x||_1 \right] + \mathbb{E}_{y \in dom(B)} \left[||Gen_A(Gen_B(y)) - y||_1 \right]$$

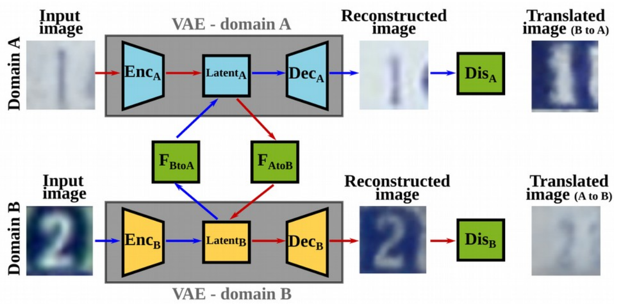
→ ensures the cycle consistency to hold.

Zhu, et al. “Unpaired image- to-image translation using cycle-consistent adversarial networks”, ICCV’17.

Proposed Method

Shape Translation

- Learn a mapping function between the [shape] domains.
  → Through an extended CycleGAN[1].



→ Applying the Loss

$$\mathcal{L} = \mathcal{L}_{GAN}(Gen_A, Dis_A) + \mathcal{L}_{GAN}(Gen_B, Dis_B) + \lambda_{cycle} * \mathcal{L}_{cyc}(Gen_A, Gen_B) + \lambda_{sim} * \boxed{\mathcal{L}_{sim}(x, y, Gen_A, Gen_B)}$$

Where:

$$\mathcal{L}_{sim}(x, y, Gen_A, Gen_B) = \mathbb{E}_{x \in dom(B)}\, d(x, Gen_A(x)) + \mathbb{E}_{y \in dom(A)}\, d(y, Gen_B(y))$$

→ favours “good” translations.

Zhu, et al. “Unpaired image- to-image translation using cycle-consistent adversarial networks”, ICCV’17.

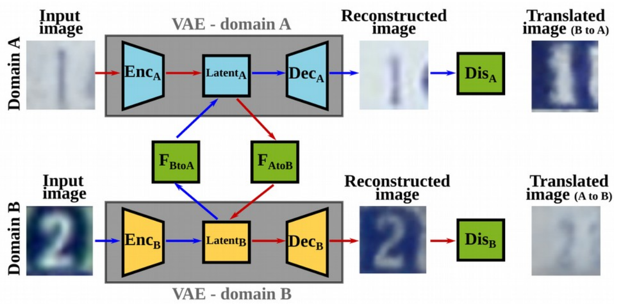
KU LEUVEN

Proposed Method

Shape Translation

- Learn a mapping function between the [shape] domains.

→ Through an extended CycleGAN[1].



→ Applying the Loss

$$\mathcal{L} = \mathcal{L}_{GAN}(Gen_A, Dis_A) + \mathcal{L}_{GAN}(Gen_B, Dis_B) + \lambda_{cycle} * \mathcal{L}_{cyc}(Gen_A, Gen_B) + \lambda_{sim} * \boxed{\mathcal{L}_{sim}(x, y, Gen_A, Gen_B)}$$

Where:

$$\mathcal{L}^{SSIM}(x,y) = \frac{1}{N}\sum_{p=1}^{N}(1 - SSIM(x_p, y_p))$$ ← perceptual similarity[2]

→ favours "good" translations.

[1]Zhu, et al. "Unpaired image- to-image translation using cycle-consistent adversarial networks", ICCV'17.
[2]Wang et al., "Multiscale structural similarity for image quality assessment," Conf.on Signals, Systems Computers, 2003.

KU LEUVEN

Evaluation

Evaluation

Translating digits in synthetic images

Color MNIST dataset[1]

~ 6.5K images per digit class.

- Results

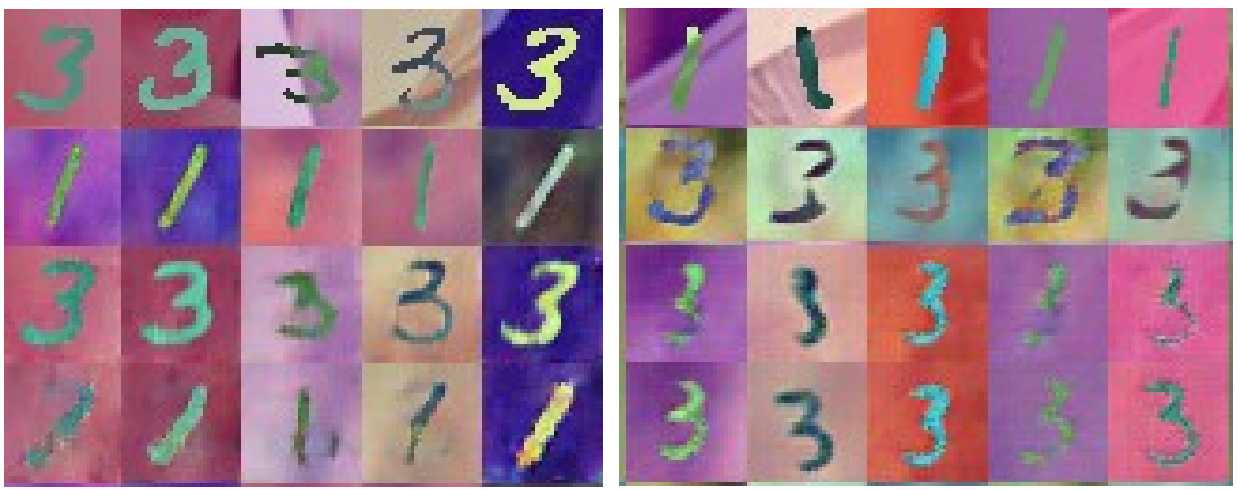
From ‘3’ to ‘1’ From ‘1’ to ‘3’

Input

Color hist. distance

DSSIM (lambda = 20)

DSSIM (lambda = 5)



Observations

- The Color loss fails at preserving the style
- The DSSIM loss tends produce blurry results for lambda=20.

[1]https://www.wouterbulten.nl/blog/tech/getting-started-with-gans-2-colorful-mnist/

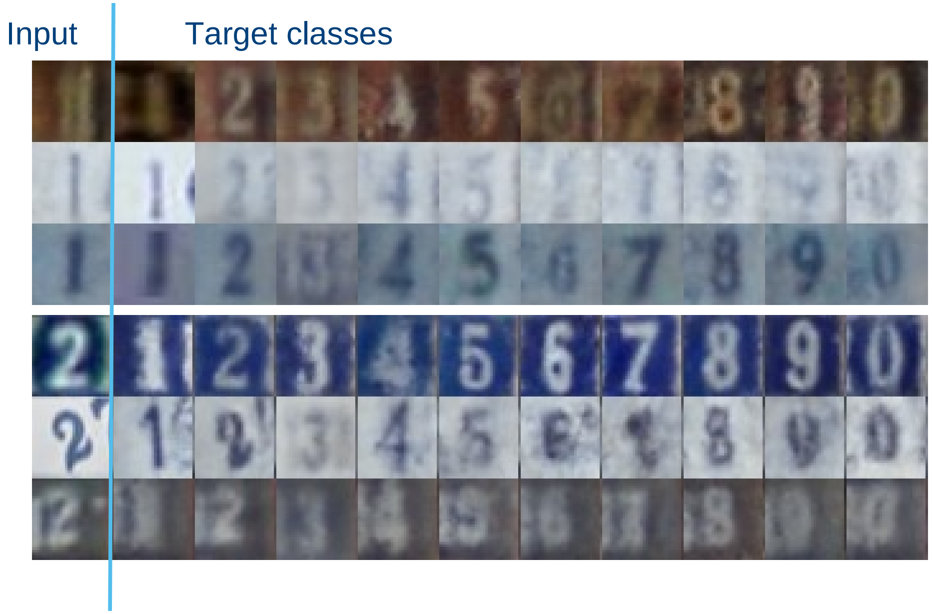
Evaluation

Translating digits in real images

SVHN dataset[1]

~ 7.3K images per digit class.

- Qualitative Results



Observations

- Overall translation is good.
- For some classes, translation is slightly blurry. (e.g. class ‘0’ & ‘6’)

[1]Coates et al., “Reading digits in natural images with unsupervised feature learning”, NIPS'11 Workshops.

Evaluation

Translating digits in real images

SVHN dataset[1]

~ 7.3K images per digit class.

- Quantitative Results

Use classification performance as a proxy metric

target class	(a)	(b)	(c)	(d)
0	96.39	85.09	65.95	75.68
1	97.94	98.47	97.71	92.50
2	95.61	95.64	92.74	95.81
3	92.68	95.91	93.29	95.40
4	96.23	96.04	93.80	91.97
5	94.80	89.77	79.58	84.62
6	95.70	88.67	74.25	81.49
7	94.06	94.95	84.37	89.47
8	93.07	86.57	73.41	82.28
9	94.98	89.34	75.09	85.06
Avg.	**95.46**	**93.47**	**86.51**	**89.27**

Classification performance from:

a) Original SVHN test set.

b) Images reconstructed by the VAEs.

c) Translated Images from class ‘1’.

d) Translated Images from class ‘2’.

[1]Coates et al., “Reading digits in natural images with unsupervised feature learning”, NIPS'11 Workshops.

Evaluation

Translating digits in real images

SVHN dataset[1]

~ 7.3K images per digit class.

- Quantitative Results

Use classification performance as a proxy metric

target class	(a)	(b)	(c)	(d)
0	96.39	85.09	65.95	75.68
1	97.94	98.47	97.71	92.50
2	95.61	95.64	92.74	95.81
3	92.68	95.91	93.29	95.40
4	96.23	96.04	93.80	91.97
5	94.80	89.77	79.58	84.62
6	95.70	88.67	74.25	81.49
7	94.06	94.95	84.37	89.47
8	93.07	86.57	73.41	82.28
9	94.98	89.34	75.09	85.06
Avg.	**95.46**	**93.47**	**86.51**	**89.27**

Classification performance from:
a) Original SVHN test set.
b) Images reconstructed by the VAEs.
c) Translated Images from class ‘1’.
d) Translated Images from class ‘2’.

Observations

- Performance is not uniform over the classes.

[1]Coates et al., “Reading digits in natural images with unsupervised feature learning”, NIPS'11 Workshops.

Evaluation

Translating digits in real images

SVHN dataset[1]

~ 7.3K images per digit class.

- Quantitative Results

Use classification performance as a proxy metric

target class	(a)	(b)	(c)	(d)
0	96.39	85.09	65.95	75.68
1	97.94	98.47	97.71	92.50
2	95.61	95.64	92.74	95.81
3	92.68	95.91	93.29	95.40
4	96.23	96.04	93.80	91.97
5	94.80	89.77	79.58	84.62
6	95.70	88.67	74.25	81.49
7	94.06	94.95	84.37	89.47
8	93.07	86.57	73.41	82.28
9	94.98	89.34	75.09	85.06
Avg.	**95.46**	**93.47**	**86.51**	**89.27**

Classification performance from:

a) Original SVHN test set.
b) Images reconstructed by the VAEs.
c) Translated Images from class ‘1’.
d) Translated Images from class ‘2’.

Observations

- Performance is not uniform over the classes.
- In some cases it is as good as on the original images.

[1]Coates et al., “Reading digits in natural images with unsupervised feature learning”, NIPS'11 Workshops.

Evaluation

Translating digits in real images

SVHN dataset[1]

~ 7.3K images per digit class.

- Quantitative Results

Use classification performance as a proxy metric

target class	(a)	(b)	(c)	(d)
0	96.39	85.09	65.95	75.68
1	97.94	98.47	97.71	92.50
2	95.61	95.64	92.74	95.81
3	92.68	95.91	93.29	95.40
4	96.23	96.04	93.80	91.97
5	94.80	89.77	79.58	84.62
6	95.70	88.67	74.25	81.49
7	94.06	94.95	84.37	89.47
8	93.07	86.57	73.41	82.28
9	94.98	89.34	75.09	85.06
Avg.	**95.46**	**93.47**	**86.51**	**89.27**

Classification performance from:

a) Original SVHN test set.
b) Images reconstructed by the VAEs.
c) Translated Images from class ‘1’.
d) Translated Images from class ‘2’.

Observations

- Performance is not uniform over the classes.
- In some cases it is as good as on the original images.
- Low translations performance seems to be correlated with weak domain modelling.

[1]Coates et al., “Reading digits in natural images with unsupervised feature learning”, NIPS'11 Workshops.

KU LEUVEN

Evaluation

Translating digits in real images

SVHN dataset[1]

~ 7.3K images per digit class.

- Quantitative Results

Use classification performance as a proxy metric

target class	(a)	(b)	(c)	(d)
0	96.39	85.09	65.95	75.68
1	97.94	98.47	97.71	92.50
2	95.61	95.64	92.74	95.81
3	92.68	95.91	93.29	95.40
4	96.23	96.04	93.80	91.97
5	94.80	89.77	79.58	84.62
6	95.70	88.67	74.25	81.49
7	94.06	94.95	84.37	89.47
8	93.07	86.57	73.41	82.28
9	94.98	89.34	75.09	85.06
Avg.	**95.46**	**93.47**	**86.51**	**89.27**

Classification performance from:

a) Original SVHN test set.

b) Images reconstructed by the VAEs.

c) Translated Images from class '1'.

d) Translated Images from class '2'.

Observations

- Performance is not uniform over the classes.
- In some cases it is as good as on the original images.
- Low translations performance seems to be correlated with weak domain modelling.

[1]Coates et al., "Reading digits in natural images with unsupervised feature learning", NIPS'11 Workshops.

Take-home Message

- **Transferring the structure while preserving the style is possible** → within simple scenarios.

- **Background clutter poses a challenge on this type of translation** → further experimentation is required.

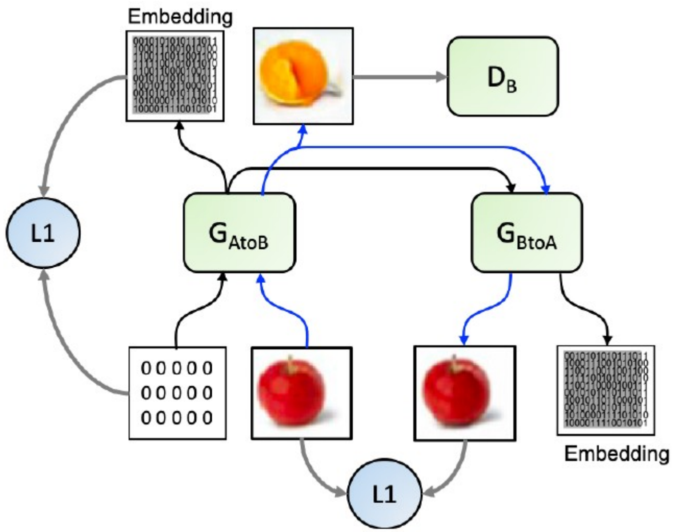
KU LEUVEN

There is Hope

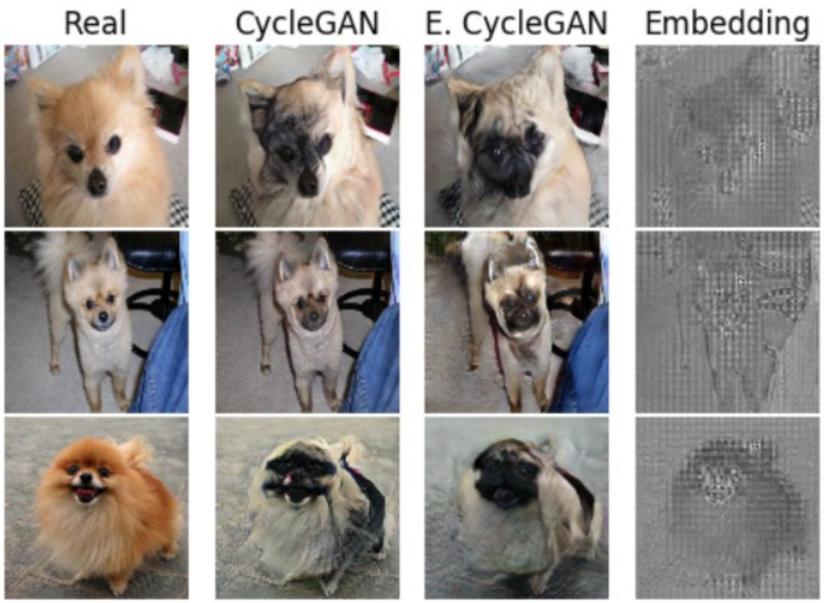
Here at ICIP'19 ! ...

Embedded CycleGAN for Shape-agnostic Translation[1]

Proposed Model



Some Results



[1] Longman & Ptucha., "Embedded cyclegan for shape-agnostic image-to-image translation", ICIP'19.

KU LEUVEN

Follow-up Work ...

Unpaired Shape Translation[1]

Clothing translation

Face translation

[1] Wang et al., “Unsupervised shape transformer for image translation and cross-domain retrieval”. ArXiv:12812.02134 .

KU LEUVEN

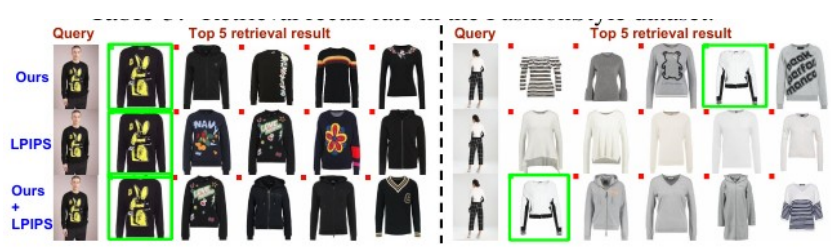
Follow-up Work ...

Unpaired Shape Translation[1]

Clothing translation

Face translation

Cross-domain → image retrieval



[1] Wang et al., “Unsupervised shape transformer for image translation and cross-domain retrieval”. arXiv:12812.02134 .

KU LEUVEN

Contact

José Oramas M.
Email: jose.oramas@esat.kuleuven.be
Website: http://homes.esat.kuleuven.be/~joramas
Twitter: @jaom7

Questions?

Towards Object Shape Translation Through Unsupervised Generative Deep Models

Lies Bollens, Tinne Tuytelaars & José Oramas M.

KU Leuven, ESAT-PSI

September 25th, 2019