Image-to-Image Translation

Jan Kautz, VP of Learning and Perception Research
This image is generated by one of our methods.
EXAMPLE USE CASES

- Low-res to high-res
- Blurry to sharp
- Image to painting
- Semantic to real
- Synthetic to real
- Thermal to color
- Day to night
- Summer to winter
- Noisy to clean
GENERATIVE ADVERSARIAL NETWORKS (GANS)

\[ p(G(z)) \rightarrow p(X) \]

[Goodfellow et al. 2014]
GAN EXAMPLE

[Karras et al. 2019]
GAN

Results look great

Not conditioned on an image

→ Condition on a given image
CONDITIONAL GANS
SUPERVISED VS UNSUPERVISED

Supervised

\[ x_i, \quad y_i \]

Unsupervised

\[ X_1, \quad X_2 \]
**UNIMODAL VS MULTIMODAL**

| Unimodal | $p(Y|X) = \delta(F(X))$ |
|----------|-------------------------|
| $F(\quad)$ | $\quad$ |

| Multimodal | $p(Y|X) = F(X, S)$ |
|------------|---------------------|
| $F(\quad)$ | $\quad$ | $\quad$ | $\quad$ | $\quad$ |
# OVERVIEW

<table>
<thead>
<tr>
<th></th>
<th>Supervised</th>
<th>Unsupervised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unimodal</td>
<td></td>
<td>UNIT</td>
</tr>
<tr>
<td>Multimodal</td>
<td>pix2pixHD, vid2vid</td>
<td>MUNIT</td>
</tr>
</tbody>
</table>
pix2pixHD: *Supervised* and *Multimodal* Image Domain Transfer

"High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs"
Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, Bryan Catanzaro
CVPR 2018

https://github.com/NVIDIA/pix2pixHD
pix2pixHD
pix2pix

\[ p(G(S), S) \rightarrow p(X, S) \]

[Isola et al. 2017]
After training
• Extract instance-pooled feature maps
• Perform k-means clustering (to features belonging to the same semantic class)

During inference
• Pick a cluster center for multimodal image translation
Coarse-to-fine Generator

Multi-scale Discriminators

Robust Objective
(GAN + discriminator feature matching loss)
pix2pixHD multimodal results
pix2pixHD label changes
vid2vid: Video-to-Video Synthesis

"Video-to-Video Synthesis"
Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Guilin Liu, Andrew Tao, Jan Kautz, Bryan Catanzaro
NIPS 2018

https://github.com/NVIDIA/vid2vid
MOTIVATION
Sequential Generator

Spatio-temporally Progressive Training

Multi-scale Discriminators

Image Discriminator

Video Discriminator

Spatially progressive

Temporally progressive

Alternating training
VID2VID

[Video-to-Video Synthesis, NeurIPS 2018]
MAKING IT INTERACTIVE

User control → Graphics engine → AI rendering → Display
AI RENDERED GAME
UNIT: Unsupervised and Unimodal Image Domain Transfer

"Unsupervised Image-to-image Translation Networks"
Ming-Yu Liu, Thomas Breuel, Jan Kautz
NIPS 2017

https://github.com/mingyuliutw/UNIT
SUPERVISED VS UNSUPERVISED

Supervised

$\ldots$

\[ x_i, \quad y_i \]

\[ \ldots \]

Unsupervised

$\ldots$

\[ X_1, \quad X_2 \]

\[ \ldots \]
\[ p(G_1(X_1) | X_1) \rightarrow p(X_2) \]

but \[ p(G_1(X_1) | X_1) \nRightarrow p(X_2 | X_1) \]

\[ p(G_2(X_2) | X_2) \rightarrow p(X_1) \]

but \[ p(G_2(X_2) | X_2) \nRightarrow p(X_1 | X_2) \]
UNIT ASSUMPTION: SHARED LATENT SPACE

What We Want

Domain 1
Encoder 1
Decoder 1
Domain 1

Domain 2
Encoder 2
Decoder 2
Domain 2

shared latent space
UNIT ASSUMPTION: SHARED LATENT SPACE

What We Get Without Constraints

Domain 1

Encoder 1

Encoder 2

shared latent space

Decoder 1

Decoder 2

Domain 2
UNIT ASSUMPTION: SHARED LATENT SPACE

Our Solution

Domain 1

Encoder 1

Encoder 2

Domain 2

Decoder 1

Decoder 2

shared latent space
DAY TO NIGHT TRANSLATION
SNOWY TO SUMMERY TRANSLATION
SUNNY TO RAINY TRANSLATION
MUNIT: Unsupervised and multimodal image domain transfer

"Multimodal Unsupervised Image-to-image Translation"
Xun Huang, Ming-Yu Liu, Serge Belongie, Jan Kautz
ECCV 2018

https://github.com/NVlabs/MUNIT
MUNIT
Learn and Control Style

Input

Sample Translations

Yosemite summer → winter

Yosemite winter → summer

[MUNIT, ECCV 2018]
MUNIT ASSUMPTION: PARTIALLY SHARED LATENT SPACE

Auto-encoding

Translation
MUNIT RESULTS
Examples with GT available

<table>
<thead>
<tr>
<th>Input</th>
<th>GT</th>
<th>Sample Translations</th>
<th>Input</th>
<th>GT</th>
<th>Sample Translations</th>
</tr>
</thead>
</table>

NVIDIA
MUNIT RESULTS

(a) Cityscape $\rightarrow$ SYNTHIA

(b) SYNTHIA $\rightarrow$ Cityscape
MUNIT RESULTS

Input

Sample Translations

(c) summer → winter

(d) winter → summer
CONCLUSION

pix2pixHD / vid2vid

(M)UNIT
YOUR LIFE’S WORK STARTS HERE

JOIN NVIDIA

100 Best Companies to Work For
  – Fortune

Most Innovative Companies
  – Fast Company

World’s Most Admired Companies
  – Fortune

Employees’ Choice: Highest Rated CEOs
  – Glassdoor

50 Smartest Companies
  – MIT Tech Review

INTERESTED? Email: aijobs@nvidia.com
THANK YOU!

@jankautz