Deep Learning for Computer Vision

Spring 2019

http://vllab.ee.ntu.edu.tw/dlcv.html (primary)
https://ceiba.ntu.edu.tw/1072CommE5052 (grade, etc.)

FB: DLCV Spring 2019

Yu-Chiang Frank Wang 王鈺強, Associate Professor
Dept. Electrical Engineering, National Taiwan University

2019/05/01
What Will Be Covered in Today’s Lecture?

• Transfer Learning & Representation Disentanglement
  • TL/RD for Visual Classification
  • TL/RD for Visual Analysis
Deep Transfer Learning

- Use of deep features
  - No transfer is performed.
  - Expect deep features are sufficient in describing cross-domain data.
  - Might not hold in practice.
Deep Transfer Learning

• Naïve transfer
  • Labeled data are available in both source and target domains
  • Train the DL model using source-domain labeled data
  • Fine-tune the MLP or few layers using target-domain labeled data
  • However, the amount of labeled data in the target domain is typically limited.
  • Possible to encounter overfitting problems
Domain Confusion by Domain-Adversarial Training

- Domain-Adversarial Training of Neural Networks (DANN)
  - Y. Ganin et al., ICML 2015
  - Maximize domain confusion = maximize domain classification loss
  - Minimize source-domain data classification loss

\[
\frac{\partial L_y}{\partial \theta_y} - \lambda \frac{\partial L_d}{\partial \theta_f}
\]

\[
d_i \log \hat{d}_i + (1 - d_i) \log(1 - \hat{d}_i)
\]
Beyond Domain Confusion

• **Domain Separation Network (DSN)**
  - Bousmalis et al., NIPS 2016
  - Separate encoders for domain-invariant and domain-specific features
  - Private/common features are *disentangled* from each other.
Beyond Domain Confusion

• Domain Separation Network, NIPS 2016
  • Example results

Original image $X_T$

Reconstruct private feature $D(E_p(x_T))$

Reconstruct private + shared features $D(E_c(x_T)+E_p(x_T))$

Reconstruct shared feature only $D(E_c(x_T))$
What Will Be Covered in Today’s Lecture?

• Transfer Learning & Representation Disentanglement
  • TL/RD for Visual Classification
    • Semantic Segmentation
    • Multi-Label Classification
    • Person Re-Identification
Semantic Segmentation Across Cities

- No More Discrimination: Cross City Adaptation of Road Scene Segmenters
  - Wang et al., ICCV 2017
  - Weakly supervised DA for semantic segmentation
Semantic Segmentation Across Cities

- No More Discrimination: Cross City Adaptation of Road Scene Segmenters
  - Wang et al., ICCV 2017
  - Weakly supervised DA for semantic segmentation
Semantic Segmentation Across Cities

• No More Discrimination: Cross City Adaptation of Road Scene Segmenters

How to get class labels in target domain?

Class-wise Domain difference
Semantic Segmentation Across Cities

• No More Discrimination: Cross City Adaptation of Road Scene Segmenters
  • Chen et al., ICCV 2017
  • Weakly supervised DA for semantic segmentation
  • Static-object prior from Google Map Time Machine features
Semantic Segmentation Across Cities

- No More Discrimination: Cross City Adaptation of Road Scene Segmenters

![Diagram of semantic segmentation across cities with feature extractor, label predictor, and domain adversarial learning.](Image)
Semantic Segmentation Across Cities

- No More Discrimination: Cross City Adaptation of Road Scene Segmenters
  - Chen et al., ICCV 2017
  - Weakly supervised DA for semantic segmentation
  - Static-object prior from Google Map Time Machine features
  - Qualitative example results
Evaluation

- 4 target-domain cities
- Evaluation metric: mean IOU (intersection over union)
Evaluation

- Visualization by t-SNE
Deep Transfer Learning for Multi-Label Classification

- Multi-label classification
  - Predicting multiple labels without observing annotated ground truth info
  - Learning across image and label-domain data + exploit label co-occurrences

Labels:
- Person
- Table
- Sofa
- Chair
- TV
- Lights
- Carpet
...

Image
Deep Transfer Learning for Multi-Label Classification

- **Our Proposed Method:** *Canonical Correlated AutoEncoder (C2AE)* [AAAI’17]
- **Highlights:**
  - Unique integration of autoencoder & deep canonical correlation analysis (DCCA)
  - Autoencoder in C2AE: label embedding + label recovery + label co-occurrence
  - DCCA in C2AE: joint feature & label embedding
  - Can handle *missing labels* during learning

Wang et al., Learning Deep Latent Spaces for Multi-Label Classification, AAAI 2017
Deep Transfer Learning for Multi-Label Classification

- Our Proposed Method: **Canonical Correlated AutoEncoder (C2AE)** [AAAI’17]

- **Highlights:**
  - Unique integration of autoencoder & deep canonical correlation analysis (DCCA)
Cross-Resolution Person Re-Identification (Re-ID)

- Person Re-ID: recognizing persons across different cameras
  - Applications: video surveillance, computational forensics, etc.

Cross-Resolution Person Re-Identification (Re-ID)

- **Person Re-ID**: recognizing persons across different cameras
  - Applications: video surveillance, computational forensics, etc.
- **Cross-Resolution Person Re-ID**
  - In practice, resolution of query images might vary.

Our Propose Deep Neural Network Architecture

• Cross-Resolution Person Re-ID

Experiments on Cross-Resolution Person Re-ID

- Quantitative Evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>MLR-CUHK03</th>
<th></th>
<th></th>
<th></th>
<th>MLR-VIPeR</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank 1</td>
<td>Rank 5</td>
<td>Rank 10</td>
<td>Rank 20</td>
<td>Rank 1</td>
<td>Rank 5</td>
<td>Rank 10</td>
<td>Rank 20</td>
</tr>
<tr>
<td>JUDEA (Li et al. 2015)</td>
<td>26.2</td>
<td>58.0</td>
<td>73.4</td>
<td>87.0</td>
<td>26.0</td>
<td>55.1</td>
<td>69.2</td>
<td>82.3</td>
</tr>
<tr>
<td>SLD²L (Jing et al. 2015)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>20.3</td>
<td>44.0</td>
<td>62.0</td>
<td>78.2</td>
</tr>
<tr>
<td>SDF (Wang et al. 2016)</td>
<td>22.2</td>
<td>48.0</td>
<td>64.0</td>
<td>80.0</td>
<td>9.25</td>
<td>38.1</td>
<td>52.4</td>
<td>68.0</td>
</tr>
<tr>
<td>SING (Jiao et al. 2018)</td>
<td>67.7</td>
<td>90.7</td>
<td>94.7</td>
<td>97.4</td>
<td>33.5</td>
<td>57.0</td>
<td>66.5</td>
<td>76.6</td>
</tr>
<tr>
<td>CSR-GAN (Wang et al. 2018b)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>37.2</td>
<td>62.3</td>
<td>71.6</td>
<td>83.7</td>
</tr>
<tr>
<td>Baseline (train on HR)</td>
<td>60.6</td>
<td>89.4</td>
<td>95.0</td>
<td>98.4</td>
<td>32.5</td>
<td>59.2</td>
<td>69.0</td>
<td>76.2</td>
</tr>
<tr>
<td>Baseline (train on HR &amp; LR)</td>
<td>65.9</td>
<td>92.1</td>
<td>97.4</td>
<td>98.9</td>
<td>36.6</td>
<td>62.3</td>
<td>70.9</td>
<td>82.2</td>
</tr>
<tr>
<td>Ours (single-level)</td>
<td>77.6</td>
<td>96.2</td>
<td>98.5</td>
<td>99.3</td>
<td>41.2</td>
<td>66.3</td>
<td>75.6</td>
<td>87.1</td>
</tr>
<tr>
<td>Ours (multi-level)</td>
<td>78.9</td>
<td>97.3</td>
<td>98.7</td>
<td>99.5</td>
<td>42.5</td>
<td>68.3</td>
<td>79.6</td>
<td>88.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>CAVIAR</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank 1</td>
<td>Rank 5</td>
<td>Rank 10</td>
<td>Rank 20</td>
</tr>
<tr>
<td>JUDEA (Li et al. 2015)</td>
<td>22.0</td>
<td>60.1</td>
<td>80.8</td>
<td>98.1</td>
</tr>
<tr>
<td>SLD²L (Jing et al. 2015)</td>
<td>18.4</td>
<td>44.8</td>
<td>61.2</td>
<td>83.6</td>
</tr>
<tr>
<td>SDF (Wang et al. 2016)</td>
<td>14.3</td>
<td>37.5</td>
<td>62.5</td>
<td>95.2</td>
</tr>
<tr>
<td>SING (Jiao et al. 2018)</td>
<td>33.5</td>
<td>72.7</td>
<td>89.0</td>
<td>98.6</td>
</tr>
<tr>
<td>CSR-GAN (Wang et al. 2018b)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Baseline (train on HR)</td>
<td>27.5</td>
<td>63.2</td>
<td>79.3</td>
<td>92.2</td>
</tr>
<tr>
<td>Baseline (train on HR &amp; LR)</td>
<td>31.7</td>
<td>68.4</td>
<td>84.2</td>
<td>94.1</td>
</tr>
<tr>
<td>Ours (single-level)</td>
<td>41.5</td>
<td>75.3</td>
<td>85.6</td>
<td>95.8</td>
</tr>
<tr>
<td>Ours (multi-level)</td>
<td>42.0</td>
<td>77.3</td>
<td>89.6</td>
<td>98.7</td>
</tr>
</tbody>
</table>

Experiments on Cross-Resolution Person Re-ID

• Qualitative Evaluation

(a) Colorization with respect to identity.  
(b) Colorization with respect to resolution.

Experiments on Cross-Resolution Person Re-ID

• Example re-ID results

What Will Be Covered in Today’s Lecture?

• Transfer Learning & Representation Disentanglement
  • TL/RD for Visual Analysis & Manipulation
    • Image Translation
    • Style Transfer
From Image Understanding to Image Manipulation

• Representation Disentanglement
  • InfoGAN
  • AC-GAN

• Image Translation
  • Pix2pix (CVPR’17)
  • CycleGAN/DualGAN/DiscoGAN
  • UNIT (NIPS’17)
  • DTN (ICLR’17)

• Joint Image Translation & Disentanglement
  • StarGAN (CVPR’18)
  • CDRD (CVPR’18)
From Image Understanding to Image Manipulation

• Image Translation
  • Pix2pix (CVPR’17): **Pairwise cross-domain training data**
  • CycleGAN/DualGAN/DiscoGAN (2017): **Unpaired cross-domain training data**
  • RecycleGAN (ECCV’18): **Unpaired cross-domain training data**
  • UNIT (NIPS’17): **Learning cross-domain image representation (with unpaired training data)**
  • BicycleGAN (NIPS’17): **Multi-modal image-to-image translation**
  • DRIT (ECCV’18): **Multi-modal image-to-image translation**

• Joint Image Translation & Disentanglement
  • StarGAN (CVPR’18): **Image translation via representation disentanglement**
  • CDRD (CVPR’18): **Cross-domain representation disentanglement and translation**
  • UFDN (NIPS’18): **Multi-domain representation disentanglement and translation**
Pix2pix

• Image-to-image translation with conditional adversarial networks (CVPR’17)
  • Can be viewed as image style transfer

Pix2pix

• **Goal / Problem Setting**
  • Image translation across two distinct domains (e.g., sketch v.s. photo)
  • Pairwise training data

• **Method: Conditional GAN**
  • Example: Sketch to Photo
    • **Generator (VAE)**
      Input: Sketch
      Output: Photo
    • **Discriminator**
      Input: Concatenation of Input(Sketch) & Synthesized/Real(Photo) images
      Output: Real or Fake

Pix2pix

- **Learning the model**

  **Training Phase**

  Overall objective function

  \[ G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \mathcal{L}_{L1}(G) \]

  **Conditional GAN loss**

  \[ \mathcal{L}_{cGAN}(G, D) = \mathbb{E}_x [\log (1 - D(x, G(x)))] + \mathbb{E}_{x, y} [\log D(x, y)] \]

  **Reconstruction Loss**

  \[ \mathcal{L}_{L1}(G) = \mathbb{E}_{x, y} [\|y - G(x)\|_1] \]

Pix2pix

• **Experiment results**

 Demo page: [https://affinelayer.com/pixsrv/](https://affinelayer.com/pixsrv/)

From Image Understanding to Image Manipulation

• Image Translation
  • Pix2pix (CVPR’17): Pairwise cross-domain training data
  • CycleGAN/DualGAN/DiscoGAN (2017): Unpaired cross-domain training data
  • RecycleGAN (ECCV’18): Unpaired cross-domain training data
  • UNIT (NIPS’17): Learning cross-domain image representation (with unpaired training data)
  • BicycleGAN (NIPS’17): Multi-modal image-to-image translation
  • DRIT (ECCV’18): Multi-modal image-to-image translation

• Joint Image Translation & Disentanglement
  • StarGAN (CVPR’18): Image translation via representation disentanglement
  • CDRD (CVPR’18): Cross-domain representation disentanglement and translation
CycleGAN/DiscoGAN/DualGAN

- **CycleGAN (CVPR’17)**
  - Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks -to-image translation with conditional adversarial networks

Paired

\[ x_i, y_i \]

\[
\begin{array}{c}
\text{\{shoes\}}, \\
\text{\{boots\}}, \\
\text{\{} \cdot \\
\end{array}
\]

1-to-1 Correspondence

Unpaired

\[ X, Y \]

\[
\begin{array}{c}
\text{\{cityscape\}}, \\
\text{\{landscape\}}, \\
\text{\{} \cdot \\
\end{array}
\]

- Easier to collect training data
- More practical

CycleGAN

- **Goal / Problem Setting**
  - Image translation across two distinct domains
  - **Unpaired** training data

- **Idea**
  - Autoencoding-like image translation
  - **Cycle consistency** between two domains

CycleGAN

• Method (Example: Photo & Painting)

• Based on 2 GANs
  • First GAN (G1, D1): Photo to Painting
  • Second GAN (G2, D2): Painting to Photo

• Cycle Consistency
  • Photo consistency
  • Painting consistency

CycleGAN

- **Method** (Example: Photo vs. Painting)
  - Based on 2 GANs
    - First GAN (G1, D1): Photo to Painting
    - Second GAN (G2, D2): Photo to Painting

- **Cycle Consistency**
  - **Photo** consistency
  - **Painting** consistency

CycleGAN

• Learning

Overall objective function

\[ G_1^*, G_2^* = \arg \min_{G_1, G_2} \max_{D_1, D_2} \mathcal{L}_{GAN}(G_1, D_1) + \mathcal{L}_{GAN}(G_2, D_2) + \mathcal{L}_{cyc}(G_1, G_2) \]

First GAN

Second GAN

• Adversarial Loss

  • First GAN (G1, D1):

\[ \mathcal{L}_{GAN}(G_1, D_1) = \mathbb{E}[\log(1 - D_1(G_1(x)))] + \mathbb{E}[\log D_1(y)] \]

  • Second GAN (G2, D2):

\[ \mathcal{L}_{GAN}(G_2, D_2) = \mathbb{E}[\log(1 - D_2(G_2(y)))] + \mathbb{E}[\log D_2(x)] \]
CycleGAN

• Learning

Overall objective function

\[ G_1^*, G_2^* = \arg \min_{G_1, G_2} \max_{D_1, D_2} \mathcal{L}_{GAN}(G_1, D_1) + \mathcal{L}_{GAN}(G_2, D_2) + \mathcal{L}_{cyc}(G_1, G_2) \]

• Consistency Loss
  • Photo and Painting consistency

\[ \mathcal{L}_{cyc}(G_1, G_2) = \mathbb{E} \left[ \| G_2(G_1(x)) - x \|_1 \right] + \left[ \| G_1(G_2(y)) - y \|_1 \right] \]
CycleGAN

- Example results

Project Page: https://junyanz.github.io/CycleGAN/
Image Translation Using Unpaired Training Data

- CycleGAN, DiscoGAN, and DualGAN

Kim et al. "Learning to Discover Cross-Domain Relations with Generative Adversarial Networks." ICML 2017
From Image Understanding to Image Manipulation

- **Image Translation**
  - Pix2pix (CVPR’17): Pairwise cross-domain training data
  - CycleGAN/DualGAN/DiscoGAN (2017): Unpaired cross-domain training data
  - RecycleGAN (ECCV’18): Unpaired cross-domain training data
  - UNIT (NIPS’17): Learning cross-domain image representation (with unpaired training data)
  - BicycleGAN (NIPS’17): Multi-modal image-to-image translation
  - DRIT (ECCV’18): Multi-modal image-to-image translation

- **Joint Image Translation & Disentanglement**
  - StarGAN (CVPR’18): Image translation via representation disentanglement
  - CDRD (CVPR’18): Cross-domain representation disentanglement and translation
Limitation of Cycle-GAN

1. Perceptual Mode Collapse
2. Bad Local Minima

• Solution $\rightarrow$ *Exploit spatiotemporal constraints*

Bansal *et al*., Recycle-GAN: Unsupervised Video Retargeting, ECCV 2018
Spatiotemporal Cycle Consistency

- Avoid Mode collapse

Bansal et al., Recycle-GAN: Unsupervised Video Retargeting, ECCV 2018
Spatiotemporal Cycle Consistency

- Get better local minima

Bansal et al., Recycle-GAN: Unsupervised Video Retargeting, ECCV 2018
Recycle-GAN

- Recycle-GAN: Introduce Predictors $P_X$, $P_Y$ and Recycle loss
Recycle-GAN

(a). Pix2Pix

Paired data

(b). Cycle-GAN

Cycle-consistency

(c). Recycle-GAN

Recycle-consistency

Bansal et al., Recycle-GAN: Unsupervised Video Retargeting, ECCV 2018
Recycle-GAN Loss

• Recurrent Loss

\[ L_r(P_X) = \sum_t ||x_{t+1} - P_X(x_{1:t})||^2, \]

• Recycle Loss

\[ L_r(G_X, G_Y, P_Y) = \sum_t ||x_{t+1} - G_X(P_Y(G_Y(x_{1:t})))||^2, \]

• Adversarial Loss

\[ \min_{G_Y} \max_{D_Y} L_g(G_Y, D_Y) = \sum_s \log D_Y(y_s) + \sum_t \log(1 - D_Y(G_Y(x_t))), \]
From Image Understanding to Image Manipulation

• Image Translation
  • Pix2pix (CVPR’17): Pairwise cross-domain training data
  • CycleGAN/DualGAN/DiscoGAN (2017): Unpaired cross-domain training data
  • RecycleGAN (ECCV’18): Unpaired cross-domain training data
  • UNIT (NIPS’17): Learning cross-domain image representation (with unpaired training data)
  • BicycleGAN (NIPS’17): Multi-modal image-to-image translation
  • DRIT (ECCV’18) : Multi-modal image-to-image translation

• Joint Image Translation & Disentanglement
  • StarGAN (CVPR’18) : Image translation via representation disentanglement
  • CDRD (CVPR’18) : Cross-domain representation disentanglement and translation
UNIT

- Unsupervised Image-to-Image Translation Networks (NIPS’17)
  - Image translation via learning cross-domain joint representation

**Stage1: Encode to the joint space**

\[ z \]: Joint latent space

\[ X_1 \rightarrow z \rightarrow X_2 \]

- Day
- Night

**Stage2: Generate cross-domain images**

\[ z \]: Joint latent space

\[ X_1 \rightarrow z \rightarrow X_2 \]

- Day
- Night

Liu et al., "Unsupervised image-to-image translation networks.", NIPS 2017
UNIT

- **Goal/Problem Setting**
  - Image translation across two distinct domains
  - **Unpaired** training image data

- **Idea**
  - Based on two parallel VAE-GAN models

Liu et al., "Unsupervised image-to-image translation networks.", NIPS 2017
UNIT

• Goal/Problem Setting
  • Image translation across two distinct domains
  • Unpaired training image data

• Idea
  • Based on two parallel VAE-GAN models
  • Learning of joint representation across image domains

Liu et al., "Unsupervised image-to-image translation networks.", NIPS 2017
UNIT

• Goal/Problem Setting
  • Image translation across two distinct domains
  • Unpaired training image data

• Idea
  • Based on two parallel VAE-GAN models
  • Learning of joint representation across image domains
  • Generate cross-domain images from joint representation

Liu et al., "Unsupervised image-to-image translation networks.\textquotedblright, NIPS 2017
UNIT

• Learning

Overall objective function
\[ G^* = \arg \min_G \max_D L_{VAE}(E_1, G_1, E_2, G_2) + L_{GAN}(G_1, D_1, G_2, D_2) \]

Variation Autoencoder Loss
\[ L_{VAE}(E_1, G_1, E_2, G_2) = \mathbb{E}[\|G_1(E_1(x_1)) - x_1\|_2] + \mathbb{E}[KL(q_1(z)||p(z))] \]
\[ + \mathbb{E}[\|G_2(E_2(x_2)) - x_2\|_2] + \mathbb{E}[KL(q_2(z)||p(z))] \]

Adversarial Loss
\[ L_{GAN}(G_1, D_1, G_2, D_2) = \mathbb{E}[\log(1 - D_1(G_1(z)))] + \mathbb{E}[\log D_1(y_1)] \]
\[ + \mathbb{E}[\log(1 - D_2(G_2(z)))] + \mathbb{E}[\log D_2(y_2)] \]
UNIT

• Learning

Overall objective function

\[
G = \arg\min_G \max_D \mathcal{L}_{VAE}(E_1, G_1, E_2, G_2) + \mathcal{L}_{GAN}(G_1, D_1, G_2, D_2)
\]

Variation Autoencoder Loss

\[
\mathcal{L}_{VAE}(E_1, G_1, E_2, G_2) = \mathbb{E}\left[\|G_1(E_1(x_1)) - x_1\|_2\right] + \mathbb{E}[KL(q_1(z)||p(z))] \\
\mathbb{E}\left[\|G_2(E_2(x_2)) - x_2\|_2\right] + \mathbb{E}[KL(q_2(z)||p(z))]
\]

Adversarial Loss

\[
\mathcal{L}_{GAN}(G_1, D_1, G_2, D_2) = \mathbb{E}[\log(1 - D_1(G_1(z)))] + \mathbb{E}[\log D_1(y_1)] \\
\mathbb{E}[\log(1 - D_2(G_2(z)))] + \mathbb{E}[\log D_2(y_2)]
\]

Reconstruction
UNIT

• **Learning**

  Overall objective function
  \[
  G = \arg \min_G \max_D \mathcal{L}_{VAE}(E_1, G_1, E_2, G_2) + \mathcal{L}_{GAN}(G_1, D_1, G_2, D_2)
  \]

  **Variation Autoencoder Loss**
  \[
  \mathcal{L}_{VAE}(E_1, G_1, E_2, G_2) = \mathbb{E} \left[ \| G_1(E_1(x_1)) - x_1 \|_2^2 \right] + \mathbb{E} \left[ \| G_2(E_2(x_2)) - x_2 \|_2^2 \right] + \mathbb{E} [\mathcal{KL}(q_1(z)||p(z))] + \mathbb{E} [\mathcal{KL}(q_2(z)||p(z))]
  \]

  **Adversarial Loss**
  \[
  \mathcal{L}_{GAN}(G_1, D_1, G_2, D_2) = \mathbb{E} [\log(1 - D_1(G_1(z)))] + \mathbb{E} [\log D_1(y_1)] + \mathbb{E} [\log(1 - D_2(G_2(z)))] + \mathbb{E} [\log D_2(y_2)]
  \]
UNIT

- **Learning**

  Overall objective function

  \[ G = \arg \min_G \max_D L_{VAE}(E_1, G_1, E_2, G_2) + L_{GAN}(G_1, D_1, G_2, D_2) \]

  Variation Autoencoder Loss

  \[ L_{VAE}(E_1, G_1, E_2, G_2) = \mathbb{E}[\|G_1(E_1(x_1)) - x_1\|^2] + \mathbb{E}[KL(q_1(z)||p(z))] - \mathbb{E}[\|G_2(E_2(x_2)) - x_2\|^2] + \mathbb{E}[KL(q_2(z)||p(z))] \]

  Adversarial Loss

  \[ L_{GAN}(G_1, D_1, G_2, D_2) = \mathbb{E}[\log(1 - D_1(G_1(z)))] + \mathbb{E}[\log D_1(y_1)] - \mathbb{E}[\log(1 - D_2(G_2(z)))] + \mathbb{E}[\log D_2(y_2)] \]

  Generated
UNIT

• Learning

Overall objective function

\[ G = \arg \min_G \max_D L_{VAE}(E_1, G_1, E_2, G_2) + L_{GAN}(G_1, D_1, G_2, D_2) \]

Variation Autoencoder Loss

\[ L_{VAE}(E_1, G_1, E_2, G_2) = \mathbb{E}[\|G_1(E_1(x_1)) - x_1\|^2] + \mathbb{E}[KL(q_1(z)||p(z))] \]
\[ + \mathbb{E}[\|G_2(E_2(x_2)) - x_2\|^2] + \mathbb{E}[KL(q_2(z)||p(z))] \]

Adversarial Loss

\[ L_{GAN}(G_1, D_1, G_2, D_2) = \mathbb{E}[\log(1 - D_1(G_1(z)))] + \mathbb{E}[\log D_1(y_1)] \]
\[ + \mathbb{E}[\log(1 - D_2(G_2(z)))] + \mathbb{E}[\log D_2(y_2)] \]
UNIT

• Example results

Sunny → Rainy

Rainy → Sunny

Real Street-view → Synthetic Street-view

Synthetic Street-view → Real Street-view

Github Page: https://github.com/mingyuliutw/UNIT

Liu et al., "Unsupervised image-to-image translation networks.“, NIPS 2017
From Image Understanding to Image Manipulation

- **Image Translation**
  - Pix2pix (CVPR’17): Pairwise cross-domain training data
  - CycleGAN/DualGAN/DiscoGAN (2017): Unpaired cross-domain training data
  - RecycleGAN (ECCV’18): Unpaired cross-domain training data
  - UNIT (NIPS’17): Learning cross-domain image representation (with unpaired training data)
  - BicycleGAN (NIPS’17): Multi-modal image-to-image translation
  - DRIT (ECCV’18): Multi-modal image-to-image translation

- **Joint Image Translation & Disentanglement**
  - StarGAN (CVPR’18): Image translation via representation disentanglement
  - CDRD (CVPR’18): Cross-domain representation disentanglement and translation
Multi-Modal Image Translation

- Image-to-image translation: Learn the mapping between two visual domains.
- Two challenges:
  - Aligned training image pairs are difficult to collect.
  - A single input may have *multiple* outputs.
- Multi-modal image translation aims to produce *diverse* outputs from given input image.
  - BicycleGAN (NIPS’17) → Pairwise training data
  - DRIT (ECCV’18 oral) & MUNIT (ECCV’18) → Unpaired training data
    (note: The structures of DRIT and MUNIT are very similar.)
BicycleGAN

• **Toward Multimodal Image-to-Image Translation (NIPS’17)**

• **Goal / Problem Setting**
  • Producing diverse images across two distinct domains.
  • **Pairwise** training data

• **Idea**
  • Combine conditional VAE-GAN and conditional Latent Regressor GAN.
BicycleGAN - Experiment

Zhu et al., Toward Multimodal Image-to-Image Translation, NIPS 2017
DRIT

- Diverse Image-to-Image Translation via Disentangled Representations (Lee et al., ECCV’18 oral)

- Goal / Problem Setting
  - Producing diverse images across two distinct domains.
  - Unpaired training data

- Idea
  - Disentangle latent into domain-invariant feature and domain-specific feature.
  - Generate cross-domain images by swapping the latent feature from each domain.
  - Applied cross-cycle consistency
Lee et al., Diverse Image-to-Image Translation via Disentangled Representations, ECCV 2018 (oral)
Method – Main Framework

Lee et al., Diverse Image-to-Image Translation via Disentangled Representations, ECCV 2018 (oral)
Method – For Attribute Feature

- **KL loss:**
  In order to perform stochastic sampling at test time.
- **Latent regression loss:**
  To encourage invertible mapping between the image and the latent space.

Lee et al., Diverse Image-to-Image Translation via Disentangled Representations, ECCV 2018 (oral)
Method – Testing phase

(b) Testing with random attributes

(c) Testing with a given attribute

Lee et al., Diverse Image-to-Image Translation via Disentangled Representations, ECCV 2018 (oral)
Domain Transfer Networks

- Unsupervised Cross-Domain Image Generation (ICLR’17)

Goal/Problem Setting
- Image translation across two domains
- One-way only translation
- Unpaired training data

Idea
- Apply unified model to learn joint representation across domains.
Domain Transfer Networks

• Unsupervised Cross-Domain Image Generation (ICLR’17)

• **Goal/Problem Setting**
  • Image translation across two domains
  • One-way only translation
  • **Unpaired** training data

• **Idea**
  • Apply unified model to learn joint representation across domains.
  • Consistency observed in image and feature spaces

---

Taigman et al., "Unsupervised cross-domain image generation.“, ICLR 2016
Domain Transfer Networks

**Learning**

- **Unified model** to translate across domains
  
  \[ G^* = \underset{G}{\arg \min} \underset{D}{\max} L_{\text{img}}(G) + L_{\text{feat}}(G) + L_{\text{GAN}}(G, D) \]

- Consistency of feature and image space
  \[
  L_{\text{img}}(G) = \mathbb{E} \left[ \| g(f(y)) - y \|_2^2 \right]
  
  L_{\text{feat}}(G) = \mathbb{E} \left[ \| f(g(f(x))) - f(x) \|_2^2 \right]
  
- Adversarial loss
  \[
  L_{\text{GAN}}(G, D) = \mathbb{E} \left[ \log(1 - D(G(x))) \right] + \mathbb{E} \left[ \log(1 - D(G(y))) \right] + \mathbb{E} \left[ \log D(y) \right]
  \]
Domain Transfer Networks

- **Learning**
  - **Unified model** to translate across domains
    \[ G^* = \arg \min_G \max_D \mathcal{L}_{\text{img}}(G) + \mathcal{L}_{\text{feat}}(G) + \mathcal{L}_{\text{GAN}}(G, D) \]
  - **Consistency of image and feature space**
    \[ \mathcal{L}_{\text{img}}(G) = \mathbb{E} \left[ \| g(f(y)) - y \|_2 \right] \]
    \[ \mathcal{L}_{\text{feat}}(G) = \mathbb{E} \left[ \| f(g(f(x))) - f(x) \|_2 \right] \]
    \[ G = \{ f, g \} \]
  - **Adversarial loss**
    \[ \mathcal{L}_{\text{GAN}}(G, D) = \mathbb{E}[\log(1 - D(G(x))) + \mathbb{E}[\log(1 - D(G(y)))] + \mathbb{E}[\log D(y)] \]
Domain Transfer Networks

- **Learning**
  - **Unified model** to translate across domains
    
    $$G^* = \arg \min_G \max_D \mathcal{L}_{img}(G) + \mathcal{L}_{feat}(G) + \mathcal{L}_{GAN}(G, D)$$

  - Consistency of feature and image space
    
    $$\mathcal{L}_{img}(G) = \mathbb{E}\left[\|g(f(x)) - y\|_2\right]$$
    
    $$\mathcal{L}_{feat}(G) = \mathbb{E}\left[\|f(g(f(x))) - f(x)\|_2\right]$$

- **Adversarial loss**
  
  $$\mathcal{L}_{GAN}(G, D) = \mathbb{E}[\log(1 - D(G(x)))] + \mathbb{E}[\log(1 - D(G(y)))] + \mathbb{E}[\log D(y)]$$
Domain Transfer Networks

- **Learning**
  - **Unified model** to translate across domains
    \[
    G^* = \arg \min_G \max_D \mathcal{L}_{img}(G) + \mathcal{L}_{feat}(G) + \mathcal{L}_{GAN}(G, D)
    \]
  - Consistency of feature and image space
    \[
    \mathcal{L}_{img}(G) = \mathbb{E} \left[ \| g(f(y)) - y \|_2 \right]
    \]
    \[
    \mathcal{L}_{feat}(G) = \mathbb{E} \left[ \| f(g(f(x))) - f(x) \|_2 \right]
    \]
  - Adversarial loss
    \[
    \mathcal{L}_{GAN}(G, D) = \mathbb{E}[\log(1 - D(G(x))] + \mathbb{E}[\log(1 - D(G(y))] + \mathbb{E}[\log D(y)]
    \]
DTN

- Example results

**SVHN 2 MNIST**

**Photo 2 Emoji**

Taigman et al., "Unsupervised cross-domain image generation.“, ICLR 2016
From Image Understanding to Image Manipulation

• Image Translation
  • Pix2pix (CVPR’17): Pairwise cross-domain training data
  • CycleGAN/DualGAN/DiscoGAN (2017): Unpaired cross-domain training data
  • RecycleGAN (ECCV’18): Unpaired cross-domain training data
  • UNIT (NIPS’17): Learning cross-domain image representation (with unpaired training data)
  • BicycleGAN (NIPS’17): Multi-modal image-to-image translation
  • DRIT (ECCV’18): Multi-modal image-to-image translation

• Joint Image Translation & Disentanglement
  • StarGAN (CVPR’18): Image translation via representation disentanglement
  • CDRD (CVPR’18): Cross-domain representation disentanglement and translation
  • UFDN (NIPS’18): Multi-domain representation disentanglement and translation
StarGAN

- Goal
  - Unified GAN for multi-domain image-to-image translation
StarGAN

- Goal
  - Unified GAN for multi-domain image-to-image translation

Traditional Cross-Domain Models

Unified Multi-Domain Model (StarGAN)

StarGAN

- **Goal / Problem Setting**
  - **Single** image translation model across **multiple** domains
  - Unpaired training data
StarGAN

- **Goal / Problem Setting**
  - Single Image translation model across multiple domains
  - Unpaired training data

- **Idea**
  - Concatenate image and target domain label as input of generator
  - Auxiliary domain classifier on Discriminator
StarGAN

• **Goal / Problem Setting**
  - Single Image translation model across multiple domains
  - Unpaired training data

• **Idea**
  - Concatenate image and target domain label as input of Generator
  - Auxiliary [domain classifier](#) as discriminator too
StarGAN

• **Goal / Problem Setting**
  - Single Image translation model across multiple domains
  - Unpaired training data

• **Idea**
  - Concatenate image and target domain label as input of Generator
  - Auxiliary domain classifier on Discriminator
  - Cycle consistency across domains
StarGAN

- **Goal / Problem Setting**
  - Single Image translation model across multiple domains
  - Unpaired training data

- **Idea**
  - Auxiliary domain classifier as discriminator
  - Concatenate image and target domain label as input
  - Cycle consistency across domains
StarGAN

• **Learning**

**Overall objective function**

\[ G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) + \mathcal{L}_{cyc}(G) \]
StarGAN

- **Learning**

  Overall objective function
  
  \[ G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) + \mathcal{L}_{cyc}(G) \]

  Adversarial Loss

- **Adversarial Loss**

  \[ \mathcal{L}_{GAN}(G, D) = \mathbb{E}[\log(1 - D(G(x, c)))] + \mathbb{E}[\log D(y)] \]
StarGAN

• **Learning**

  Overall objective function
  \[
  G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) + \mathcal{L}_{Cyc}(G)
  \]

  Domain Classification Loss

  • Adversarial Loss
  \[
  \mathcal{L}_{GAN}(G, D) = \mathbb{E}[\log(1 - D(G(x, c)))] + \mathbb{E}[\log D(y)]
  \]

  • **Domain Classification Loss (Disentanglement)**
  \[
  \mathcal{L}_{cls}(G, D) = \mathbb{E}[-\log D_{cls}(c'|y)] + \mathbb{E}[-\log D_{cls}(c|G(x, c))]
  \]
StarGAN

• **Learning**

Overall objective function

\[ G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) + \mathcal{L}_{Cyc}(G) \]

Domain Classification Loss

• **Adversarial Loss**

\[ \mathcal{L}_{GAN}(G, D) = \mathbb{E}[\log(1 - D(G(x, c)))] + \mathbb{E}[\log D(y)] \]

• **Domain Classification Loss (Disentanglement)**

\[ \mathcal{L}_{cls}(G, D) = \mathbb{E}[- \log D_{cls}(c'|y)] + \mathbb{E}[- \log D_{cls}(c|G(x, c))] \]

Real data w.r.t. its domain label
StarGAN

**Learning**

Overall objective function

\[ G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) + \mathcal{L}_{Cyc}(G) \]

- **Adversarial Loss**

\[ \mathcal{L}_{GAN}(G, D) = \mathbb{E}[\log(1 - D(G(x, c)))] \]

- **Domain Classification Loss (Disentanglement)**

\[ \mathcal{L}_{cls}(G, D) = \mathbb{E}[- \log D_{cls}(c^'|y)] + \mathbb{E}[- \log D_{cls}(c|G(x, c))] \]

Generated data w.r.t. assigned label
StarGAN

- **Learning**

  Overall objective function
  
  \[ G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G,D) + \mathcal{L}_{CLS}(G,D) + \mathcal{L}_{Cyc}(G) \]

  Consistency Loss

  - **Adversarial Loss**
    \[ \mathcal{L}_{GAN}(G,D) = \mathbb{E}[\log(1 - D(G(x,c)))] + \mathbb{E}[\log D(y)] \]

  - **Domain Classification Loss (Disentanglement)**
    \[ \mathcal{L}_{CLS}(G,D) = \mathbb{E}[- \log D_{cls}(c'|y)] + \mathbb{E}[- \log D_{cls}^{-1}(c|G(x,c))] \]

  - **Cycle Consistency Loss**
    \[ \mathcal{L}_{Cyc}(G) = \mathbb{E}[\|G(G(x,c),c_x) - x\|_1] \]
StarGAN

• **Learning**

**Overall objective function**

\[ G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) + \mathcal{L}_{cyc}(G) \]

• **Adversarial Loss**

\[ \mathcal{L}_{GAN}(G, D) = \mathbb{E}[\log(1 - D(G(x, c)))] + \mathbb{E}[\log D(y)] \]

• **Domain Classification Loss**

\[ \mathcal{L}_{cls}(G, D) = \mathbb{E}[-\log D_{cls}(c'|y)] + \mathbb{E}[-\log D_{cls}(c|G(x, c))] \]

• **Cycle Consistency Loss**

\[ \mathcal{L}_{cyc}(G) = \mathbb{E}[\|G(G(x, c), c_x) - x\|_1] \]
StarGAN

• Example results
  • StarGAN can somehow be viewed as a representation disentanglement model, instead of an image translation one.

Multiple Domains

Github Page: https://github.com/yunjey/StarGAN

From Image Understanding to Image Manipulation

• Image Translation
  • Pix2pix (CVPR’17): Pairwise cross-domain training data
  • CycleGAN/DualGAN/DiscoGAN (2017): Unpaired cross-domain training data
  • RecycleGAN (ECCV’18): Unpaired cross-domain training data
  • UNIT (NIPS’17): Learning cross-domain image representation (with unpaired training data)
  • BicycleGAN (NIPS’17): Multi-modal image-to-image translation
  • DRIT (ECCV’18): Multi-modal image-to-image translation

• Joint Image Translation & Disentanglement
  • StarGAN (CVPR’18): Image translation via representation disentanglement
  • CDRD (CVPR’18): Cross-domain representation disentanglement and translation
  • UFDN (NIPS’18): Multi-domain representation disentanglement and translation
Cross-Domain Representation Disentanglement (CDRD)

- **Goal / Problem Setting**
  - Learning cross-domain joint disentangled representation
  - Single domain supervision
  - Cross-domain image translation with attribute of interest

- **Idea**
  - Bridge the domain gap across domains
  - Auxiliary attribute classifier on Discriminator
  - Semantics of disentangled factor is learned from label in source domain
**CDRD**

**Goal / Problem Setting**
- Learning cross-domain joint disentangled representation
- Single domain supervision
- Cross-domain image translation with attribute of interest

**Idea**
- Bridge the domain gap across domains
- Auxiliary attribute classifier on Discriminator
- Semantics of disentangled factor is learned from label in source domain
CDRD

**Goal / Problem Setting**
- Learning cross-domain joint disentangled representation
- Single domain supervision
- Cross-domain image translation with attribute of interest

**Idea**
- Bridge the domain gap across domains
- Auxiliary attribute classifier on Discriminator
- Semantics of disentangled factor is learned from label in source domain
CDRD

**Goal / Problem Setting**
- Learning cross-domain joint disentangled representation
- Single domain supervision
- Cross-domain image translation with attribute of interest

**Idea**
- Based on GAN
- Auxiliary attribute classifier on Discriminator
- Bridge the domain gap across domains
- Semantics of disentangled factor is learned from label in source domain
CDRD

• **Goal / Problem Setting**
  - Learning cross-domain joint disentangled representation
  - Single domain supervision
  - Cross-domain image translation with attribute of interest

• **Idea**
  - Based on GAN
  - Auxiliary attribute classifier as Discriminator
  - Bridge the domain gap across domains
  - Semantics of disentangled factor is learned from label in source domain
CDRD

- **Goal / Problem Setting**
  - Learning cross-domain joint disentangled representation
  - Single domain supervision
  - Cross-domain image translation with attribute of interest

- **Idea**
  - Based on GAN
  - Auxiliary attribute classifier on Discriminator
  - Bridge the domain gap with division of high and low-level layers
  - Semantics of disentangled factor is learned from label in source domain
CDRD

- **Goal / Problem Setting**
  - Learning cross-domain joint disentangled representation
  - Single domain supervision
  - Cross-domain image translation with attribute of interest

- **Idea**
  - Based on GAN
  - Auxiliary attribute classifier on Discriminator
  - Bridge the domain gap with division of high- and low-level layer
  - **Semantics** of disentangled factor is learned from label info in source domain
CDRD

Learning

Overall objective function

\[ G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) \]
CDRD

• Learning

Overall objective function

\[ G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) \]

Adversarial Loss

\[ \mathcal{L}_{GAN}(G, D) = \mathcal{L}_{GAN}^S(G, D) + \mathcal{L}_{GAN}^T(G, D) \]

\[ \mathcal{L}_{GAN}^S(G, D) = \mathbb{E}[\log(D_C(D_S(X_S)))] + \mathbb{E}[\log(1 - D_C(D_S(\tilde{X}_S)))] \]

\[ \mathcal{L}_{GAN}^T(G, D) = \mathbb{E}[\log(D_C(D_T(X_T)))] + \mathbb{E}[\log(1 - D_C(D_T(\tilde{X}_T)))] \]
CDRD

• **Learning**

  Overall objective function

  \[
  G^* = \arg\min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D)
  \]

  Adversarial Loss

  \[
  \mathcal{L}_{GAN}(G, D) = \mathcal{L}^S_{GAN}(G, D) + \mathcal{L}^T_{GAN}(G, D)
  \]

  \[
  \mathcal{L}^S_{GAN}(G, D) = \mathbb{E}[\log(D_C(D_S(X_S)))] + \mathbb{E}[\log(1 - D_C(D_S(\tilde{X}_S)))]
  \]

  \[
  \mathcal{L}^T_{GAN}(G, D) = \mathbb{E}[\log(D_C(D_T(X_T)))] + \mathbb{E}[\log(1 - D_C(D_T(\tilde{X}_T)))]
  \]
CDRD

• **Learning**

  Overall objective function
  
  \[ G^* = \arg\min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) \]

  Adversarial Loss

  \[ \mathcal{L}_{GAN}(G, D) = \mathcal{L}^S_{GAN}(G, D) + \mathcal{L}^T_{GAN}(G, D) \]

  \[ \mathcal{L}^S_{GAN}(G, D) = \mathbb{E}[\log(D_C(D_S(X_S)))] + \mathbb{E}[\log(1 - D_C(D_S(\tilde{X}_S)))] \]

  \[ \mathcal{L}^T_{GAN}(G, D) = \mathbb{E}[\log(D_C(D_T(X_T)))] + \mathbb{E}[\log(1 - D_C(D_T(\tilde{X}_T)))] \]

  **Generated**
CDRD

- **Learning**

  Overall objective function

  \[ G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D) \]

  **Adversarial Loss**

  \[
  \mathcal{L}_{GAN}(G, D) = \mathcal{L}^S_{GAN}(G, D) + \mathcal{L}^T_{GAN}(G, D)
  \]

  \[
  \mathcal{L}^S_{GAN}(G, D) = \mathbb{E} [\log(D_C(D_S(X_S)))] + \mathbb{E} [\log(1 - D_C(D_S(\tilde{X}_S)))]
  \]

  \[
  \mathcal{L}^T_{GAN}(G, D) = \mathbb{E} [\log(D_C(D_T(X_T)))] + \mathbb{E} [\log(1 - D_C(D_T(\tilde{X}_T)))]
  \]

  **Disentangle Loss**

  \[
  \mathcal{L}_{cls}(G, D) = \mathcal{L}^S_{cls}(G, D) + \mathcal{L}^T_{cls}(G, D)
  \]

  \[
  \mathcal{L}^S_{cls}(G, D) = \mathbb{E} [\log P(l = \tilde{l}|\tilde{X}_S)] + \mathbb{E} [\log P(l = l_S|X_S)] \quad \text{AC-GAN}
  \]

  \[
  \mathcal{L}^T_{cls}(G, D) = \mathbb{E} [\log P(l = \tilde{l}|\tilde{X}_T)]
  \]
**CDRD**

- **Learning**

  **Overall objective function**
  
  $$G^* = \arg \min_G \max_D \mathcal{L}_{GAN}(G, D) + \mathcal{L}_{cls}(G, D)$$

  **Adversarial Loss**
  
  $$\mathcal{L}_{GAN}(G, D) = \mathcal{L}^S_{GAN}(G, D) + \mathcal{L}^T_{GAN}(G, D)$$
  
  $$\mathcal{L}^S_{GAN}(G, D) = \mathbb{E}[\log(D_C(D_S(X_S)))] + \mathbb{E}[\log(1 - D_C(D_S(\tilde{X}_S)))]$$
  
  $$\mathcal{L}^T_{GAN}(G, D) = \mathbb{E}[\log(D_C(D_T(X_T)))] + \mathbb{E}[\log(1 - D_C(D_T(\tilde{X}_T)))]$$

  **Disentangle Loss**
  
  $$\mathcal{L}_{cls}(G, D) = \mathcal{L}^S_{cls}(G, D) + \mathcal{L}^T_{cls}(G, D)$$
  
  $$\mathcal{L}^S_{cls}(G, D) = \mathbb{E}[\log P(l = \tilde{l}|X_S)] + \mathbb{E}[\log P(l = l_S|X_S)]$$
  
  $$\mathcal{L}^T_{cls}(G, D) = \mathbb{E}[\log P(l = \tilde{l}|\tilde{X}_T)]$$ **InfoGAN**
CDRD

- Add an additional encoder
  - **Input:** Gaussian Noise Image
  - Image translation with attribute of interest
CDRD

- **Experiment results**

Wang et al., Detach and Adapt: Learning Cross-Domain Disentangled Deep Representation”. *CVPR 2018*
• Experiment results

Cross-Domain Classification

<table>
<thead>
<tr>
<th>Domain</th>
<th>( \tilde{I} )</th>
<th>NIPS '16 CoGAN</th>
<th>NIPS '17 UNIT</th>
<th>CDRD</th>
<th>E-CDRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>sketch (S)</td>
<td>smiling</td>
<td>89.50</td>
<td>90.10</td>
<td>90.19</td>
<td>90.01</td>
</tr>
<tr>
<td></td>
<td>photo (T)</td>
<td>78.90</td>
<td>81.04</td>
<td>87.61</td>
<td>88.28</td>
</tr>
<tr>
<td>sketch (S)</td>
<td>glasses</td>
<td>96.63</td>
<td>97.65</td>
<td>97.06</td>
<td>97.19</td>
</tr>
<tr>
<td></td>
<td>photo (T)</td>
<td>81.01</td>
<td>79.89</td>
<td>94.49</td>
<td>94.84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain</th>
<th>( \tilde{I} )</th>
<th>NIPS '16 CoGAN</th>
<th>NIPS '17 UNIT</th>
<th>CDRD</th>
<th>E-CDRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>photo (S)</td>
<td>night</td>
<td>98.04</td>
<td>98.49</td>
<td>97.06</td>
<td>97.14</td>
</tr>
<tr>
<td></td>
<td>paint (T)</td>
<td>65.18</td>
<td>67.81</td>
<td>84.21</td>
<td>85.58</td>
</tr>
<tr>
<td>photo (S)</td>
<td>season</td>
<td>86.74</td>
<td>85.64</td>
<td>86.21</td>
<td>88.92</td>
</tr>
<tr>
<td></td>
<td>paint (T)</td>
<td>65.94</td>
<td>66.09</td>
<td>79.87</td>
<td>80.03</td>
</tr>
</tbody>
</table>

Wang et al., Detach and Adapt: Learning Cross-Domain Disentangled Deep Representation”. CVPR 2018

Figure: t-SNE for digit
# Comparisons

<table>
<thead>
<tr>
<th></th>
<th>Cross-Domain Image Translation</th>
<th>Representation Disentanglement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unpaired Training Data</td>
<td>Multi-domains</td>
</tr>
<tr>
<td>Pix2pix</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CycleGAN</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>StarGAN</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>UNIT</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>DTN</td>
<td>O</td>
<td>X</td>
</tr>
<tr>
<td>infoGAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC-GAN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDRD (Ours)</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

*Cannot translate image across domains*
From Image Understanding to Image Manipulation

• Image Translation
  • Pix2pix (CVPR’17): Pairwise cross-domain training data
  • CycleGAN/DualGAN/DiscoGAN (2017): Unpaired cross-domain training data
  • RecycleGAN (ECCV’18): Unpaired cross-domain training data
  • UNIT (NIPS’17): Learning cross-domain image representation (with unpaired training data)
  • BicycleGAN (NIPS’17): Multi-modal image-to-image translation
  • DRIT (ECCV’18) : Multi-modal image-to-image translation

• Joint Image Translation & Disentanglement
  • StarGAN (CVPR’18) : Image translation via representation disentanglement
  • CDRD (CVPR’18) : Cross-domain representation disentanglement and translation
  • UFDN (NIPS’18) : Multi-domain representation disentanglement and translation
A Unified Feature Disentangler for Multi-Domain Image Translation and Manipulation

- Learning interpretable representations
A Unified Feature Disentangler for Multi-Domain Image Translation and Manipulation

- Learning interpretable representations
A Unified Feature Disentangler for Multi-Domain Image Translation and Manipulation

- Comparisons

<table>
<thead>
<tr>
<th>Method</th>
<th>Unpaired data</th>
<th>Bidirectional translation</th>
<th>Unified structure</th>
<th>Multiple domains</th>
<th>Joint representation</th>
<th>Feature disentanglement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pix2Pix [10]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CycleGAN [30]</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>StarGAN [4]</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>DTN [24]</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>UNIT [16]</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>-</td>
<td>✓</td>
<td>-</td>
</tr>
<tr>
<td>E-CDRD [18]</td>
<td>✓</td>
<td>✓</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>UFDN (Ours)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Example Results

• Translation of Face Images
Example Results

- Multi-attribute image translation
What We’ve Covered in Today’s Lecture?

• Transfer Learning & Representation Disentanglement
  • TL/RD for Visual Classification
  • TL/RD for Visual Analysis

• 5/8 Wed 10am
  • I’m out for ICLR @ New Orleans
  • Invited Talk by Dr. Yen-Yu Lin from Academia Sinica
  • Covers his recent works on image segmentation and video analysis
  • More details to come