Mix and match networks: encoder-decoder alignment for zero-pair image translation

Yaxing Wang, Joost Van De Weijer, Luis Herranz

Computer Vision Center, Universitat Autònoma de Barcelona, Spain

**INTRODUCTION**

**Problem:** Image translation between multiple domains/modalities. Most image translation methods (e.g., pix2pix[1], CycleGAN[2]) address one-to-one mappings.

**Zero-pair translation:** translation between domains or modalities for which no direct paired data is available during training.

**Approach:**
- **Mix and match networks:** create any translation (even unseen, i.e. zero-pair) by simply assembling encoder-decoder pairs.
- **Challenge:** robust cross-domain encoder-decoder alignment.

**Benefits:**
- Scalable many-to-many mapping
- Unique domain translation
- Effective network

**EXPERIMENTS**

Mix and match networks for object recoloring and style transfer.

**Mix and match networks for zero-shot cross-modal translation.**

Dataset (based on SceneNetRGBD [3]):
- Train: 75K pairs (RGB,depth) and 75K pairs (RGB,segmentation).
- Test (zero-pair): 10K pairs (depth, segmentation).

<table>
<thead>
<tr>
<th>Method</th>
<th>Conn.</th>
<th>λ_u</th>
<th>δ</th>
<th>RMSE (lin)</th>
<th>RMSE (log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baselines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CycleGAN [2]</td>
<td>-</td>
<td>CE</td>
<td>2.79 0.00</td>
<td>16.9 6.81</td>
<td>4.88 0.82</td>
</tr>
<tr>
<td>2×pix2pix</td>
<td>[1]</td>
<td>CE</td>
<td>34.6 1.88</td>
<td>70.9 20.9</td>
<td>63.6 17.6</td>
</tr>
<tr>
<td>M&amp;MNets</td>
<td>D → R,S</td>
<td>CE</td>
<td>0.02 0.00</td>
<td>8.76 0.30</td>
<td>2.91 2.06</td>
</tr>
<tr>
<td>M&amp;MNets</td>
<td>D → R,S</td>
<td>SC</td>
<td>25.4 0.26</td>
<td>82.7 0.44</td>
<td>56.6 6.30</td>
</tr>
<tr>
<td>M&amp;MNets</td>
<td>D ↔ S</td>
<td>PI</td>
<td>56.8 18.9</td>
<td>48.9 31.4</td>
<td>88.7 48.3</td>
</tr>
</tbody>
</table>

**Encoder-decoder alignment**

We use several techniques to enforce alignment:
- Shared encoders and decoders
- Autoencoders (AE)
- Latent consistency losses (LCL)
- Robust side-information (pooling indices)

**References**

