Few-shot font generation...
- Generate images of all characters from a few reference samples
- Outputs should have consistent style with the samples' content information (what the letter is)

A few samples

\[ \text{A} \quad \text{G} \quad \text{J} \rightarrow \quad \text{A} \quad \text{B} \quad \text{C} \quad \text{D} \quad \text{E} \quad \text{F} \quad \text{G} \quad \text{H} \quad \text{I} \quad \text{J} \ldots \]

Whole character set (=Font)

\[ \text{永} \quad \text{中} \quad \text{右} \rightarrow \quad \text{理} \quad \text{有} \quad \text{合} \quad \text{塗} \quad \text{章} \quad \text{本} \quad \text{系} \quad \text{和} \ldots \]

Related Works

1. EMD [Zhang+., CVPR18]
The architecture for B/W Chinese font

2. AGIS-Net [Gao+., TOG19]
The architecture with GAN

Problem: Low quality when the number of style reference is extremely limited

Proposed Method

Introducing **metric learning** to the style encoder

- L2-constrained softmax loss [Ranjani+, arXiv17]

![Proposed Method Diagram](image)

- Embedding space
  - Same style: bring embeddings close
  - Different style: keep embeddings apart

Style encoder can focus only on **style** characteristics, independent from the content of samples

Experiment

Backbone networks: EMD, AGIS-Net

- # of style reference samples: \( n \in \{5, 10, 15, 30\} \)

Dataset: Japanese typographic fonts
- 368 fonts (358 for training, 30 for evaluation)
- 2965 glyphs for each font
- 64 x 64 grayscale image

<table>
<thead>
<tr>
<th></th>
<th>( n = 5 )</th>
<th>( n = 10 )</th>
<th>( n = 15 )</th>
<th>( n = 20 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>0.386</td>
<td>0.554</td>
<td>0.275</td>
<td>0.607</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.274</td>
<td>0.566</td>
<td>0.219</td>
<td>0.666</td>
</tr>
<tr>
<td>AGIS-Net + DML (ours)</td>
<td>0.189</td>
<td>0.701</td>
<td>0.179</td>
<td>0.721</td>
</tr>
<tr>
<td>EMD</td>
<td>0.187</td>
<td>0.705</td>
<td>0.179</td>
<td>0.726</td>
</tr>
</tbody>
</table>

We observed remarkable improvement, especially when \( n \) is extremely limited.