Learning Separable Hidden Unit Contributions for Speaker-Adaptive Lip-Reading
Songtao Luo, Shuang Yang, Shiguang Shan, Xilin Chen

Motivation

- Speaker-dependent features are well-represented in the shallow layers. As the depth increases, the level of abstraction improves only slightly.
- Content-dependent features have relatively poorer representation in the shallow layers. As the depth increases, the level of abstraction improves uniformly.

Experiments

Dataset
We establish CAS-VSR-S68:
- 68-hour 11 hosts 3,800 Chinese characters
- LRW-ID
  - word-level English 500words
  - Speaker Adaptation Split of LRW
- GRID
  - sentence-level English fixed corpus

Ablation Study
Loss Ablation

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>87.25</td>
</tr>
<tr>
<td>Enhance Only</td>
<td>87.83</td>
</tr>
<tr>
<td>Suppress Only</td>
<td>87.81</td>
</tr>
<tr>
<td>Enhance &amp; Suppress</td>
<td>87.91</td>
</tr>
</tbody>
</table>