Supplementary Material for “Prototype Completion with Primitive Knowledge for Few-Shot Learning”

Baoquan Zhang, Xutao Li, Yunming Ye, Zhichao Huang, Lisai Zhang
Harbin Institute of Technology, Shenzhen
zhangbaoquan@stu.hit.edu.cn, {lixutao, yeyunming}@hit.edu.cn, iceshzc@stu.hit.edu.cn, LisaiZhang@foxmail.com

A. Additional Ablation Study

We conduct supplementary ablation studies on the tieredImagenet and CUB-200-2011 datasets, respectively. The results are shown in Table 1 and 2, which also demonstrate the effectiveness of the two specially designed components, i.e., learning to complete prototypes and Gaussian-based prototype fusion strategy (GaussFusion).

Table 1. Ablation study on tieredImagenet.

<table>
<thead>
<tr>
<th>LCP</th>
<th>GF</th>
<th>MF</th>
<th>5-way 1-shot</th>
<th>5-way 5-shot</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>√</td>
<td>√</td>
<td>69.02 ± 0.72%</td>
<td>79.31 ± 0.18%</td>
</tr>
<tr>
<td>(ii)</td>
<td>√</td>
<td></td>
<td>71.66 ± 0.92%</td>
<td>80.78 ± 0.75%</td>
</tr>
<tr>
<td>(iii)</td>
<td>√</td>
<td>√</td>
<td>74.02 ± 0.89%</td>
<td>83.29 ± 0.18%</td>
</tr>
<tr>
<td>(iv)</td>
<td>√</td>
<td>√</td>
<td>81.04 ± 0.89%</td>
<td>87.42 ± 0.57%</td>
</tr>
</tbody>
</table>

Table 2. Ablation study on CUB-200-2011.

<table>
<thead>
<tr>
<th>LCP</th>
<th>GF</th>
<th>MF</th>
<th>5-way 1-shot</th>
<th>5-way 5-shot</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>√</td>
<td>√</td>
<td>77.75 ± 0.82%</td>
<td>91.36 ± 0.41%</td>
</tr>
<tr>
<td>(ii)</td>
<td>√</td>
<td></td>
<td>84.36 ± 0.66%</td>
<td>89.19 ± 0.47%</td>
</tr>
<tr>
<td>(iii)</td>
<td>√</td>
<td>√</td>
<td>88.87 ± 0.58%</td>
<td>93.99 ± 0.34%</td>
</tr>
<tr>
<td>(iv)</td>
<td>√</td>
<td>√</td>
<td>93.20 ± 0.45%</td>
<td>94.90 ± 0.31%</td>
</tr>
</tbody>
</table>

B. Derivation of GaussFusion

Proposition. Let \( f(x) \) and \( g(x) \) be a Multivariate Gaussian Distributions with diagonal covariance, i.e., \( f(x) = N(\mu_k, \text{diag}(\sigma^2_k)) \) and \( g(x) = N(\mu_k, \text{diag}(\sigma^2_k)) \) where \( x \) is a \( d \)-dimension random vector, \( \hat{\mu}_k \) and \( \mu_k \) denote \( d \)-dimension mean vector, and \( \sigma^2_k \) and \( \sigma_k \) are \( d \)-dimension variance vector. Then, their product obeys a new Multivariate Gaussian Distributions \( N(\mu'_k, \text{diag}(\sigma'^2_k)) \) with \( \mu'_k = \frac{\sigma^2_k \odot \mu_k + \sigma_k^2 \odot \mu_k}{\sigma_k^2 + \sigma^2_k} \) and \( \sigma'_k = \frac{\sigma^2_k \odot \sigma^2_k}{\sigma_k^2 + \sigma^2_k} \), where \( \odot \) denotes the element-wise product.

Derivation. Considering that the covariances of \( f(x) \) and \( g(x) \) are simplified as diagonal covariances. This means that the variables of the random vector \( x \) are uncorrelated. In this case, \( f(x) \) and \( g(x) \) can be simplified as the expression below:

\[
\begin{align*}
    f(x) &= \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi \sigma^2_{k,i}}} e^{-\frac{(x_i - \hat{\mu}_{k,i})^2}{2\sigma^2_{k,i}}} \\
    g(x) &= \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi \sigma^2_{k,i}}} e^{-\frac{(x_i - \mu_{k,i})^2}{2\sigma^2_{k,i}}}
\end{align*}
\]

Thus, their product \( h(x) \) satisfies:

\[
\begin{align*}
    h(x) &= f(x)g(x) \\
    &= \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi \sigma^2_{k,i}}} e^{-\frac{(x_i - \hat{\mu}_{k,i})^2}{2\sigma^2_{k,i}}} \frac{1}{\sqrt{2\pi \sigma^2_{k,i}}} e^{-\frac{(x_i - \mu_{k,i})^2}{2\sigma^2_{k,i}}}
\end{align*}
\]

\[
\begin{align*}
    &= \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi \sigma^2_{k,i}}} e^{-\frac{(x_i - \hat{\mu}_{k,i})^2}{2\sigma^2_{k,i}}} + (x_i - \mu_{k,i})^2
\end{align*}
\]

\[
\begin{align*}
    &= \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi \sigma^2_{k,i}}} e^{-\frac{(x_i - \hat{\mu}_{k,i})^2}{2\sigma^2_{k,i}}} + (x_i - \mu_{k,i})^2
\end{align*}
\]

\[
\begin{align*}
    &= \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi \sigma^2_{k,i}}} e^{-\frac{(x_i - \hat{\mu}_{k,i})^2}{2\sigma^2_{k,i}}} + (x_i - \mu_{k,i})^2
\end{align*}
\]

\[
\begin{align*}
    &= \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi \sigma^2_{k,i}}} e^{-\frac{(x_i - \hat{\mu}_{k,i})^2}{2\sigma^2_{k,i}}} + (x_i - \mu_{k,i})^2
\end{align*}
\]

\[
\begin{align*}
    &= \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi \sigma^2_{k,i}}} e^{-\frac{(x_i - \hat{\mu}_{k,i})^2}{2\sigma^2_{k,i}}} + (x_i - \mu_{k,i})^2
\end{align*}
\]

where \( S_i = \frac{1}{\sqrt{2\pi (\sigma_{k,i}^2 + \sigma_{k,i}^2)}} e^{-\frac{(x_i - \mu_{k,i})^2}{2(\sigma_{k,i}^2 + \sigma_{k,i}^2)}}. \) Thus, \( h(x) \) is also a multivariate Gaussian distribution, i.e., \( N(\mu'_k, \text{diag}(\sigma'^2_k)) \) with mean \( \mu'_k = \frac{\sigma^2_k \odot \mu_k + \sigma_k^2 \odot \mu_k}{\sigma_k^2 + \sigma^2_k} \) and diagonal covariance \( \text{diag}(\sigma'^2_k) \) where \( \sigma'_k = \frac{\sigma^2_k \odot \sigma^2_k}{\sigma_k^2 + \sigma^2_k} \).