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.

	LCP	GF	MF	5-way 1-shot	5-way 5-shot
(i)				$69.02 \pm 0.72\%$	$79.31 \pm 0.18\%$
(ii)	√			$71.66 \pm 0.92\%$	$80.78 \pm 0.75\%$
(iii)	√		√	$74.02 \pm 0.89\%$	$83.29 \pm 0.18\%$
(iv)	√	√		$81.04 \pm 0.89\%$	$87.42 \pm 0.57\%$

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

	LCP	GF	MF	5-way 1-shot	5-way 5-shot
(i)				$77.75 \pm 0.82\%$	$91.36 \pm 0.41\%$
(ii)	√			84.36 ± 0.68 %	$89.19 \pm 0.47\%$
(iii)	√		√	$88.87 \pm 0.58\%$	$93.99 \pm 0.34\%$
(iv)	√	√		$93.20 \pm 0.45\%$	$94.90 \pm 0.31\%$

B. Derivation of GaussFusion

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

$$f(x) = \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi\hat{\sigma}_{k,i}^2}} e^{(\frac{-(x_i - \hat{\mu}_{k,i})^2}{2\hat{\sigma}_{k,i}^2})}$$

$$g(x) = \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi\sigma_{k,i}^2}} e^{(\frac{-(x_i - \mu_{k,i})^2}{2\sigma_{k,i}^2})}$$

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

$$\begin{aligned}
&h(x) \\
&= f(x)g(x) \\
&= \prod_{i=0}^{d-1} \frac{1}{\sqrt{2\pi\hat{\sigma}_{k,i}^2}} e^{(\frac{-(x_i - \hat{\mu}_{k,i})^2}{2\hat{\sigma}_{k,i}^2})} \frac{1}{\sqrt{2\pi\sigma_{k,i}^2}} e^{(\frac{-(x_i - \mu_{k,i})^2}{2\sigma_{k,i}^2})} \\
&= \prod_{i=0}^{d-1} \frac{1}{2\pi\sqrt{\hat{\sigma}_{k,i}^2\sigma_{k,i}^2}} e^{(\frac{-(x_i - \hat{\mu}_{k,i})^2}{2\hat{\sigma}_{k,i}^2} + \frac{-(x_i - \mu_{k,i})^2}{2\sigma_{k,i}^2})} \\
&= \prod_{i=0}^{d-1} \frac{1}{2\pi\sqrt{\hat{\sigma}_{k,i}^2\sigma_{k,i}^2}} e^{(\frac{(x_i - \frac{\sigma_{k,i}^2\hat{\mu}_{k,i} + \hat{\sigma}_{k,i}^2\mu_{k,i}}{\hat{\sigma}_{k,i}^2 + \sigma_{k,i}^2})^2}{2\frac{\sigma_{k,i}^2\hat{\sigma}_{k,i}^2}{\hat{\sigma}_{k,i}^2 + \sigma_{k,i}^2}} + \frac{(\hat{\mu}_{k,i} - \mu_{k,i})^2}{2(\hat{\sigma}_{k,i}^2 + \sigma_{k,i}^2)})} \\
&= \prod_{i=0}^{d-1} \frac{S_i}{\sqrt{2\pi\frac{\sigma_{k,i}^2\hat{\sigma}_{k,i}^2}{\hat{\sigma}_{k,i}^2 + \sigma_{k,i}^2}}} e^{(-\frac{-(x_i - \frac{\sigma_{k,i}^2\hat{\mu}_{k,i} + \hat{\sigma}_{k,i}^2\mu_{k,i}}{\hat{\sigma}_{k,i}^2 + \sigma_{k,i}^2})^2}{2(\frac{\sigma_{k,i}^2\hat{\sigma}_{k,i}^2}{\hat{\sigma}_{k,i}^2 + \sigma_{k,i}^2})})}
\end{aligned}$$

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

*Corresponding author