Gestalt-Guided Image Understanding For Few-Shot Learning

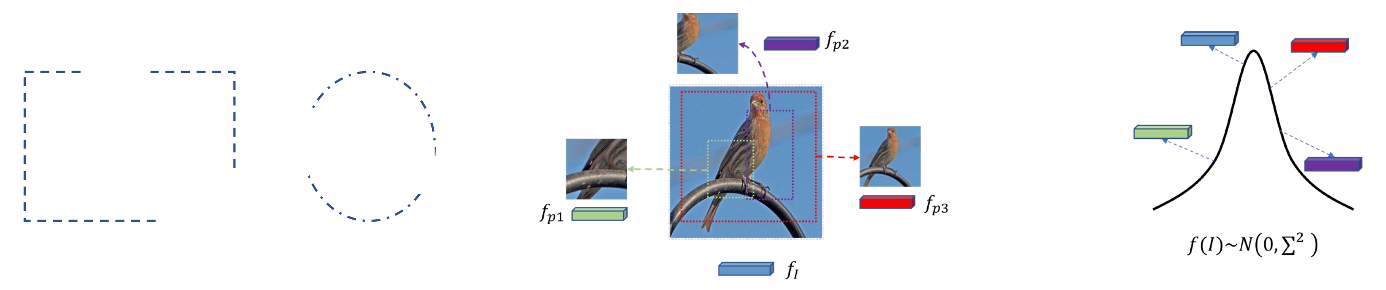
Kun Song, Yuchen Wu, Jiansheng Chen, Tianyu Hu, Huimin Ma
University of Science and Technology Beijing

ACCV 2022 Macau

Motivation:

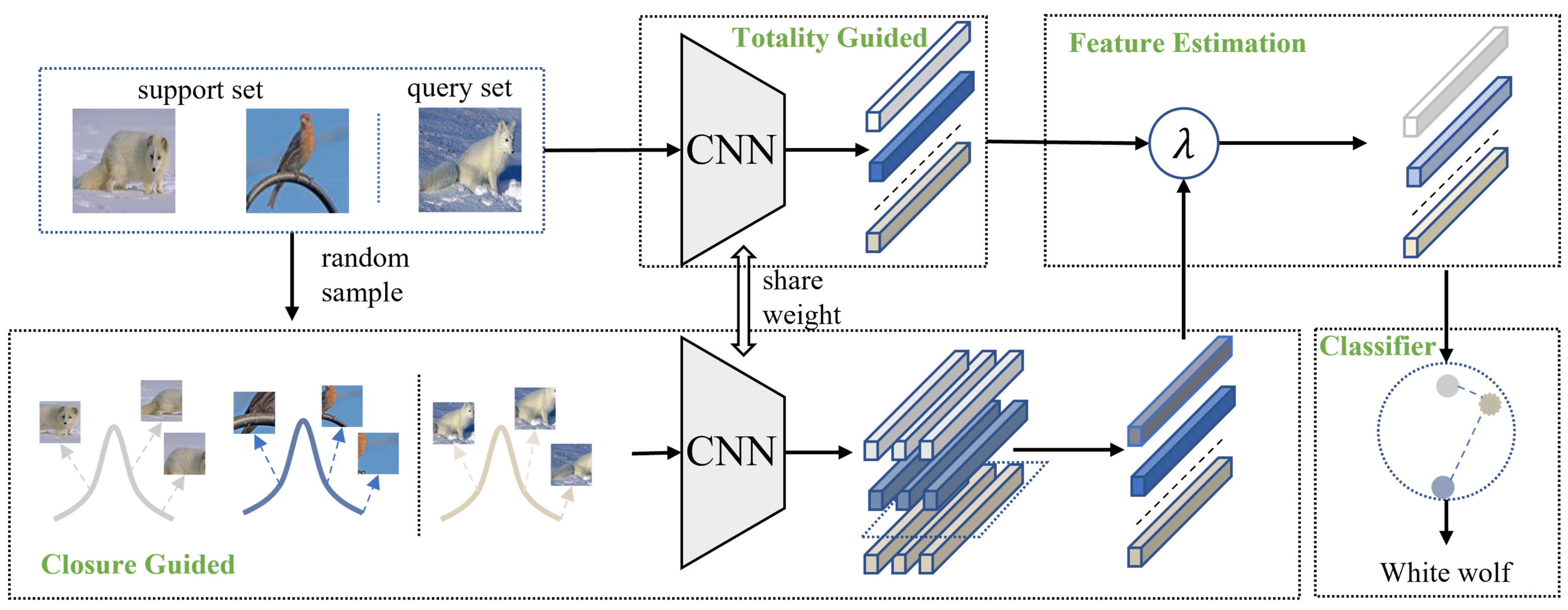
- **Gestalt psychology**

· Gestalt psychology is a psychology school that emerged in Austria and Germany in the early twentieth century. Closure is one of the laws of Gestalt psychology. Humans often consider objects as complete rather than focusing on the gaps they may have. For example, a circle has a good Gestalt in terms of completeness. we may also consider an incomplete circle or rectangle as a complete one. This tendency to complete shapes and figures is called closure. Therefore, it is easy to recognize this image is a bird. The feature of the whole image can represent the image. What's more, considering the closure. Randomly crop a patch for the image. We can still recognize that it belongs to a bird. Therefore, the feature of the patches can represent the image, too. To estimate the accurate feature of an image. We assume that the feature of the patch obeys Gaussian distribution. Mu can represent the image. The feature of image and patches are samples of the distribution. We can estimate mean from the whole image and patches, respectively.



- **Contribution**

· We introduce Gestalt psychology into the process of image understanding and propose a plug-and-play method without retraining or fine-tuning, called GGIU.
· We innovatively propose to use multivariate Gaussian distribution to describe the image and design a feature estimation module with reference to Kalman filter to estimate image feature accurately.
· We conduct extensive experiments to demonstrate the applicability of our method to a variety of different few-shot classification tasks. The experiment results demonstrate the robustness and scalability of our method.



Totality-Guided

Estimation: $\mu_t = f_I$

Error: $e_t \sim N(0, \Sigma_t^2)$

Closure-Guided

Estimation: $\mu_c = \frac{1}{M}\sum_{\{i=1\}}^{M} f_{p_i}$

Error: $e_t \sim N(0, \Sigma_c^2)$

Feature Estimation

Estimation: $f = \mu_c + \lambda(\mu_t - \mu_c)$

Error: $e \sim N(0, \Sigma_e^2)$

covariance matrix:

$$\Sigma_e = \lambda\Sigma_t\lambda^T + (1-\lambda)\Sigma_c(1-\lambda)^T$$

Minimize $tr(\Sigma_e)$

$$\frac{\partial\Sigma_e}{\partial\lambda} = 0$$

$$\lambda = (\Sigma_t\Sigma_c^{-1} + I)^{-1}$$

Method	Backbone	5way-1shot	5way-5shot
PN*	Conv-4	49.42 ± 0.78	**68.20±0.66**
PN†		50.15 ± 0.44	65.19 ± 0.51
DC*		**54.62±0.64**	-
Spot and Learn*		51.03 ± 0.78	67.96 ± 0.71
PN+GGIU	Conv-4	52.55 ± 0.52	67.36 ± 0.55
PN†	ResNet-12	61.59 ± 0.54	76.75 ± 0.46
CC*		55.45 ± 0.89	70.13 ± 0.68
CC†		63.11 ± 0.74	80.43 ± 0.31
PN+TRAML*		60.31 ± 0.48	77.94 ± 0.57
PN+CL*		59.54 ± 0.47	74.46 ± 0.52
PN+CL†		63.74 ± 0.59	79.33 ± 0.31
DC*		61.50 ± 0.47	
AA*		58.84 ± 0.77	80.35 ± 0.73
PN+GGIU†	ResNet-12	64.34 ± 0.53	79.49 ± 0.41
CC+GGIU†		**65.72±0.77**	**82.55±0.29**
PN+CL+GGIU†		65.50 ± 0.45	80.76 ± 0.39
CLIP†	ViT-B/32	88.21 ± 0.33	97.47 ± 0.08
CLIP+GGIU†	ViT-B/32	**89.31±0.33**	**97.71±0.06**

Results on miniImageNet. * represents the results reported by the original paper and † represents the results that we implement.

Method	5way-1shot	5way-5shot
PN	61.59 ± 0.54	76.75 ± 0.46
PN+GGIU	64.34 ± 0.53 (↑2.75)	79.49 ± 0.41 (↑2.74)
CC	63.11 ± 0.74	80.43 ± 0.31
CC+GGIU	65.72 ± 0.77 (↑2.61)	65.72 ± 0.77 (↑2.61)
CL	63.74 ± 0.59	79.33 ± 0.31
CL+GGIU	65.50 ± 0.45 (↑1.76)	80.76 ± 0.39 (↑1.43)
CLIP	88.21 ± 0.33	97.47 ± 0.08
CLIP+GGIU	89.31 ± 0.33 (↑1.10)	97.71 ± 0.06 (↑0.24)

Results on miniImageNet

Method	5way-1shot	5way-5shot
PN	76.13 ± 0.21	88.06 ± 0.09
PN+GGIU	78.79 ± 0.24 (↑2.66)	89.69 ± 0.17 (↑1.63)
CC	70.57 ± 0.35	86.65 ± 0.16
CC+GGIU	72.60 ± 0.29 (↑2.03)	87.90 ± 0.27 (↑1.25)
CL	72.34 ± 0.48	85.93 ± 0.25
CL+GGIU	73.64 ± 0.46 (↑1.30)	87.17 ± 0.25 (↑1.24)

Results on CUB200

Method	5way-1shot	5way-5shot
PN	40.47 ± 0.21	56.14 ± 0.20
PN+GGIU	42.61 ± 0.49 (↑2.14)	58.95 ± 0.49 (↑2.81)
CC	43.56 ± 0.47	61.51 ± 0.39
CC+GGIU	45.88 ± 0.48 (↑2.32)	64.77 ± 0.27 (↑3.26)
CL	38.65 ± 0.44	52.36 ± 0.35
CL+GGIU	39.87 ± 0.31 (↑1.22)	53.74 ± 0.21 (↑1.38)

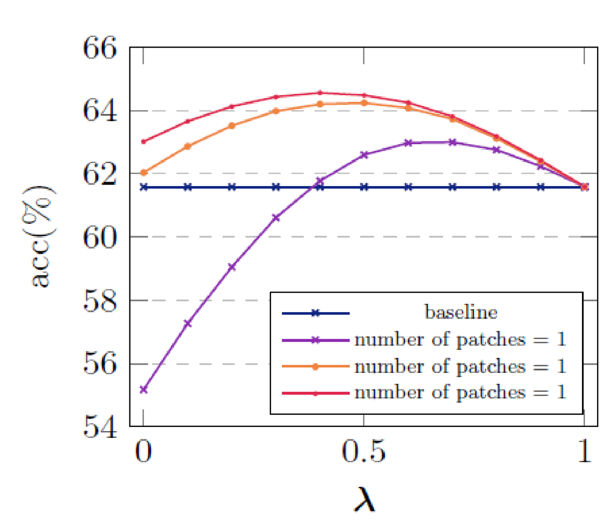
miniImageNet → CUB200

Method	5way-1shot	5way-5shot
PN	40.24 ± 0.35	55.47 ± 0.47
PN+GGIU	43.17 ± 0.57 (↑2.93)	58.12 ± 0.49 (↑2.65)
CC	43.54 ± 0.52	60.40 ± 0.39
CC+GGIU	44.27 ± 0.42 (↑0.73)	60.94 ± 0.42 (↑0.54)
CL	44.47 ± 0.56	61.84 ± 0.44
CL+GGIU	46.42 ± 0.68 (↑1.95)	63.89 ± 0.45 (↑2.05)

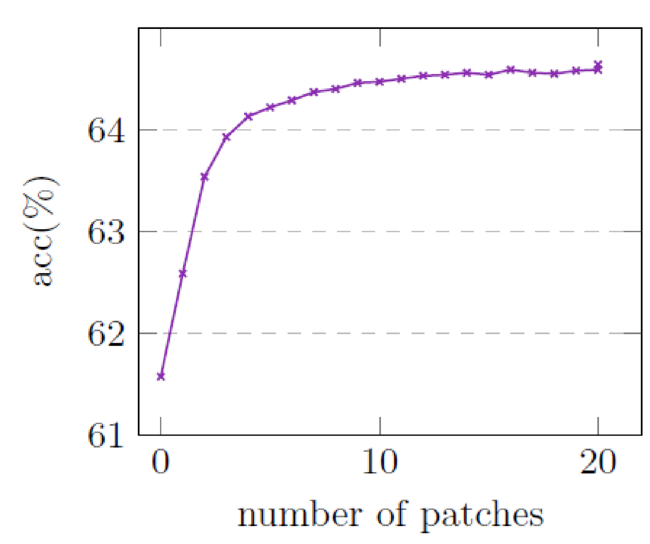
CUB200 → miniImageNet

Support	query	5way-1shot	5way-5shot
		61.59	76.75
√		62.80	77.01
	√	62.50	77.90
√	√	64.34	79.49

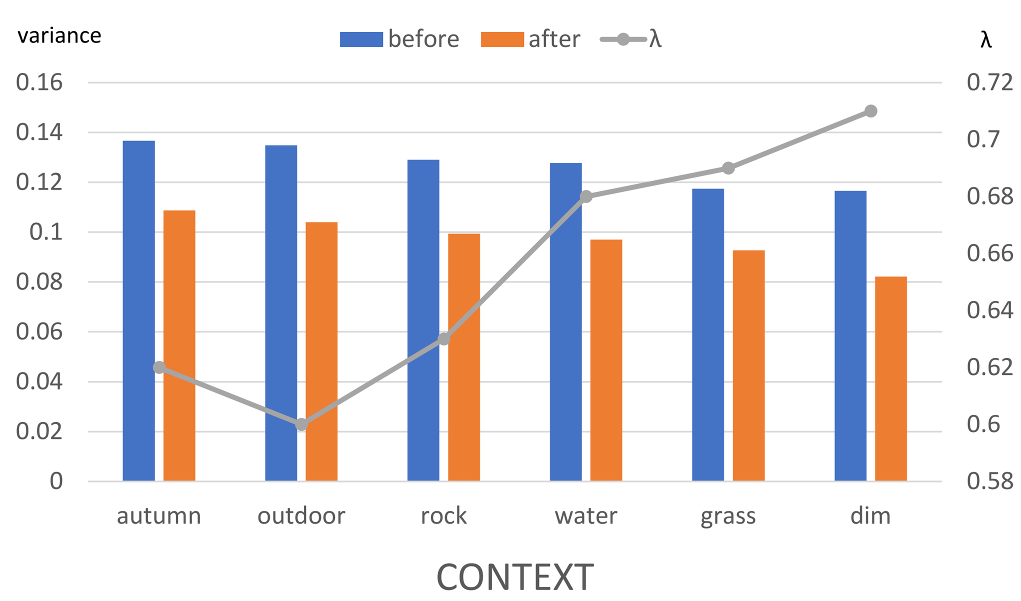
Results of ablation experiments



Relationship between λ and performance



The influence of the number of patches on performance



Relationship between intra-class variations and λ

- **Contact**
- **sonkun@xs.ustb.edu.cn**
- **Yuchen.wu@xs.ustb.edu.cn**
- **jschen@ustb.edu.cn**
- **tianyu@ustb.edu.cn**
- **mhmpub@ustb.edu.cn**