Explanation-Guided Training for Cross-Domain Few-Shot Classification

Speaker： Jiamei Sun

Jiamei Sun[1], Sebastian Lapuschkin[2], Wojciech Samek[2], Alexander Binder[1]

[1]Information System of Technology and Design, Singapore University of Technology and Design

[2]Department of Video Coding & Analytics, Fraunhofer Heinrich Hertz Institute, Berlin, Germany

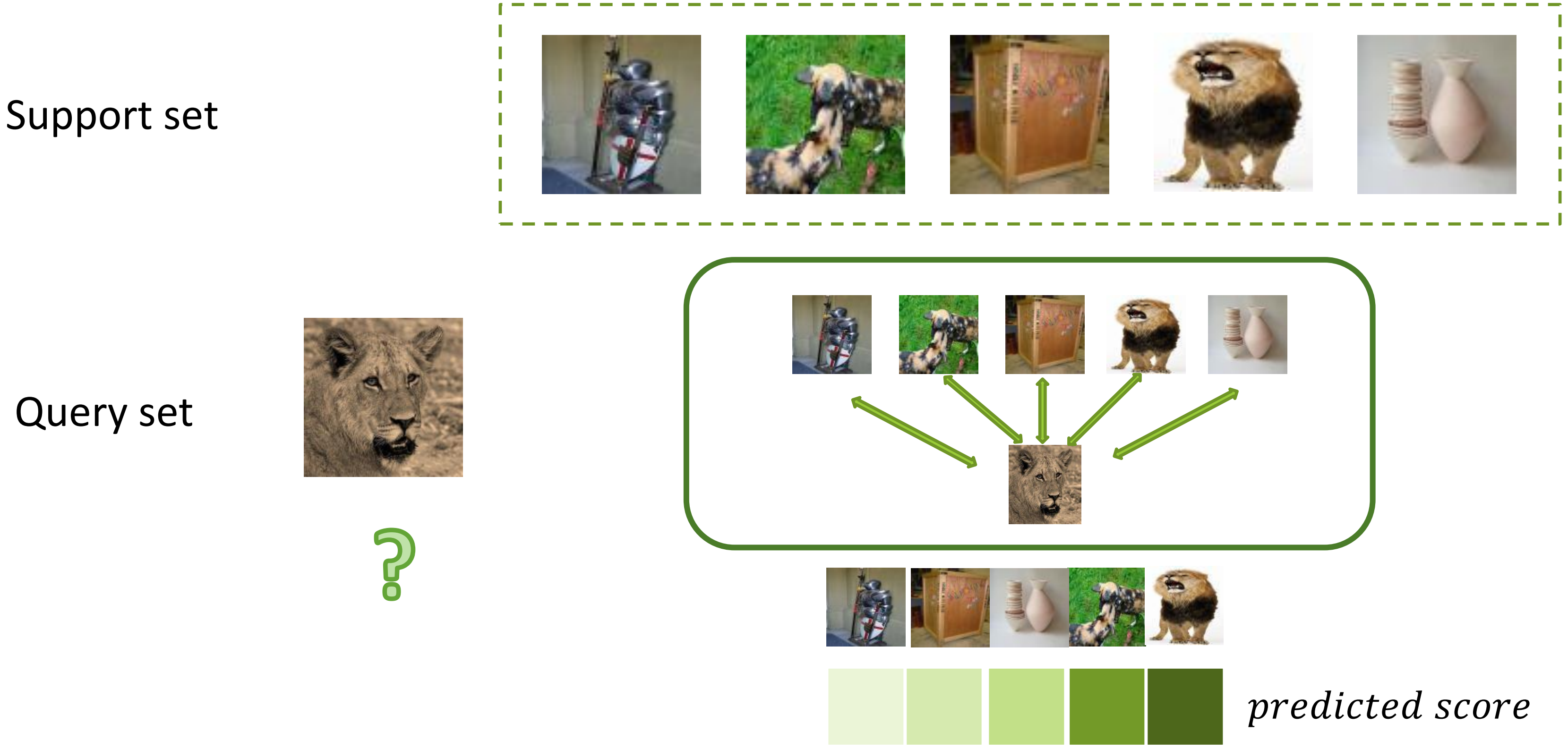
Outlines

- Few-shot classification and the challenges of ***cross-domain*** few-shot classification.
- **Interpreting** few-shot classification models with **LRP**.
- **Explanation-guided training** for metric-based few-shot classification
- Performance and effects
- Conclusion

Outlines

- Few-shot classification and the challenges of ***cross-domain*** few-shot classification.
- Interpreting few-shot classification models with LRP.
- Explanation-guided training for metric-based few-shot classification
- Performance and effects
- Conclusion

Few-shot Classification models

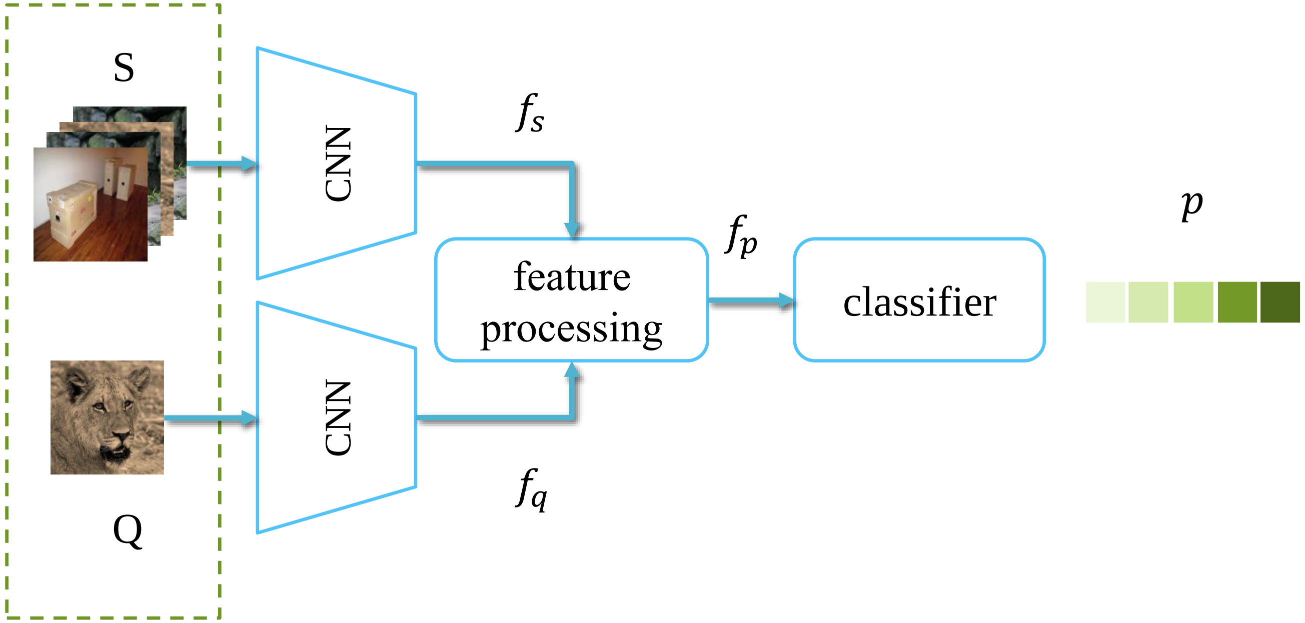


Support set

Query set

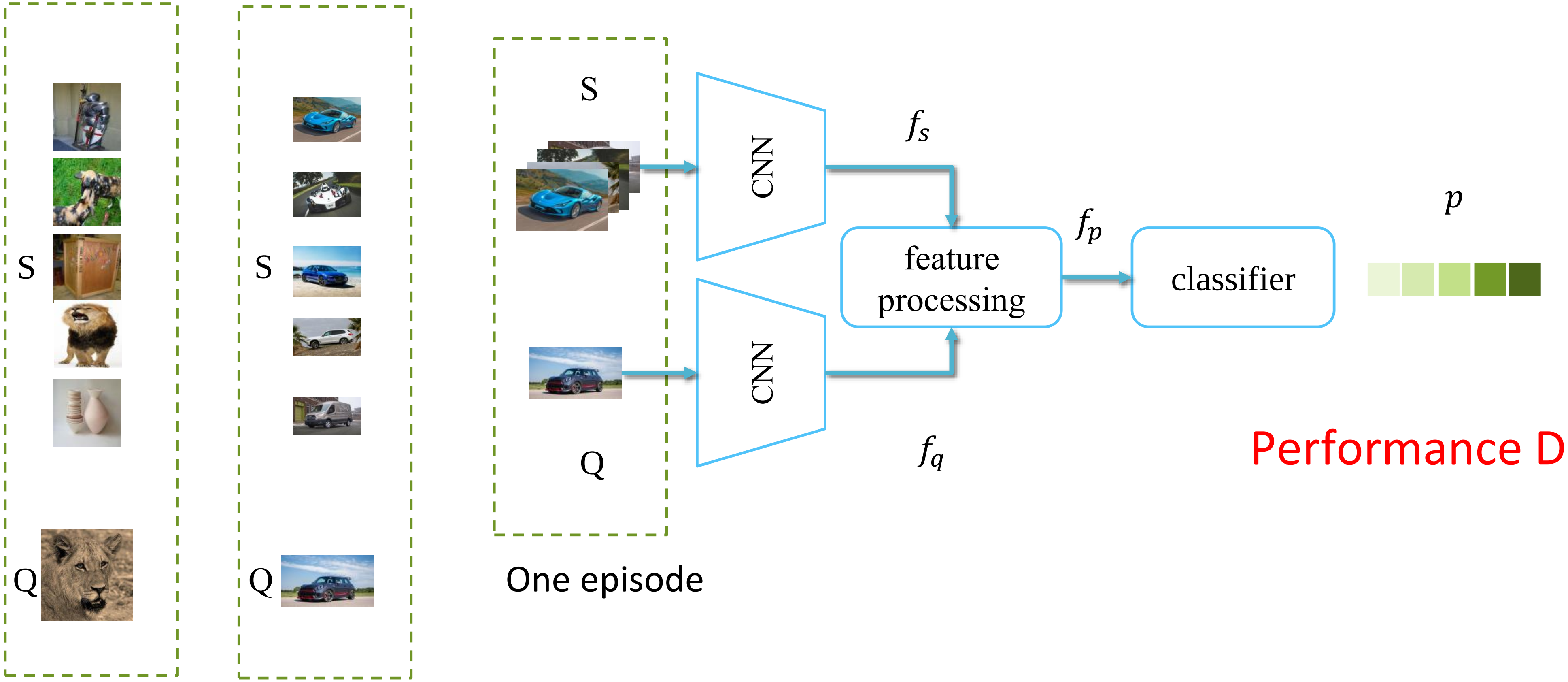
predicted score

Few-shot Classification models



One episode

Few-shot Classification models



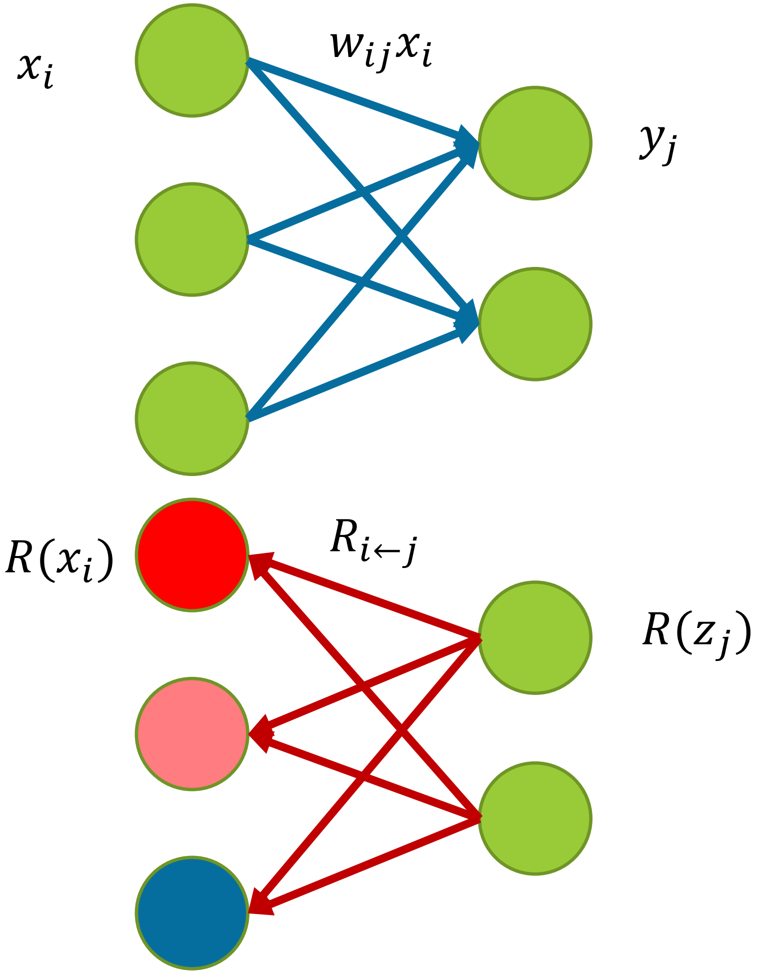
Performance Drops

miniImageNet → Cars

Outlines

- Few-shot classification and the challenges of ***cross-domain*** few-shot classification.
- **Interpreting** few-shot classification models with **LRP**.
- Explanation-guided training for metric-based few-shot classification
- Performance and effects
- Conclusion

Layer-wise Relevance Propagation



$$y_j = w_{ij}x_i + b_j$$

$$z_j = f(y_j)$$

🔴 **+** support

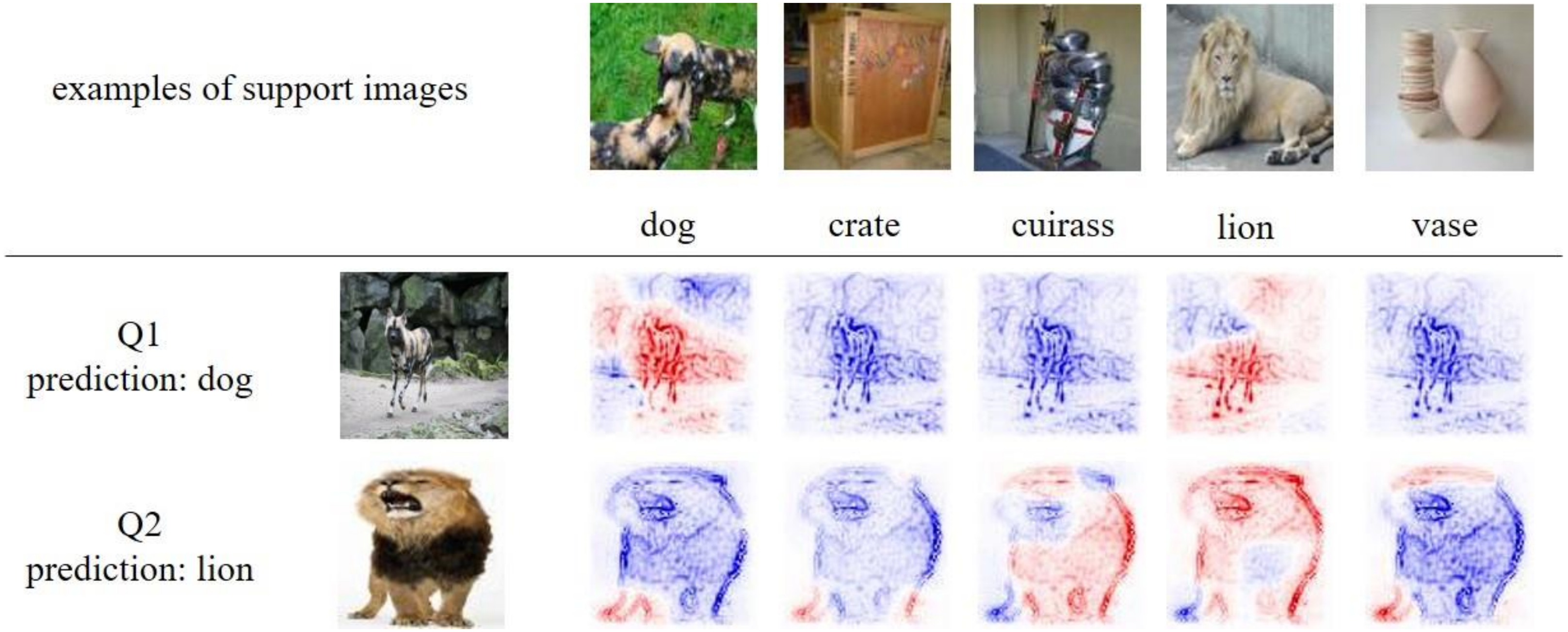
🔵 **-** opposition

$$R_{i\leftarrow j} = \frac{x_i w_{ij}}{y_j + \epsilon \odot sign(y_j)} R(z_j)$$

$$R_{i\leftarrow j} = \left(\alpha \frac{(x_i w_{ij})^{+}}{y_j^{+}} - (\alpha - 1)\frac{(x_i w_{ij})^{-}}{y_j^{-}}\right)$$

Layer-wise Relevance Propagation

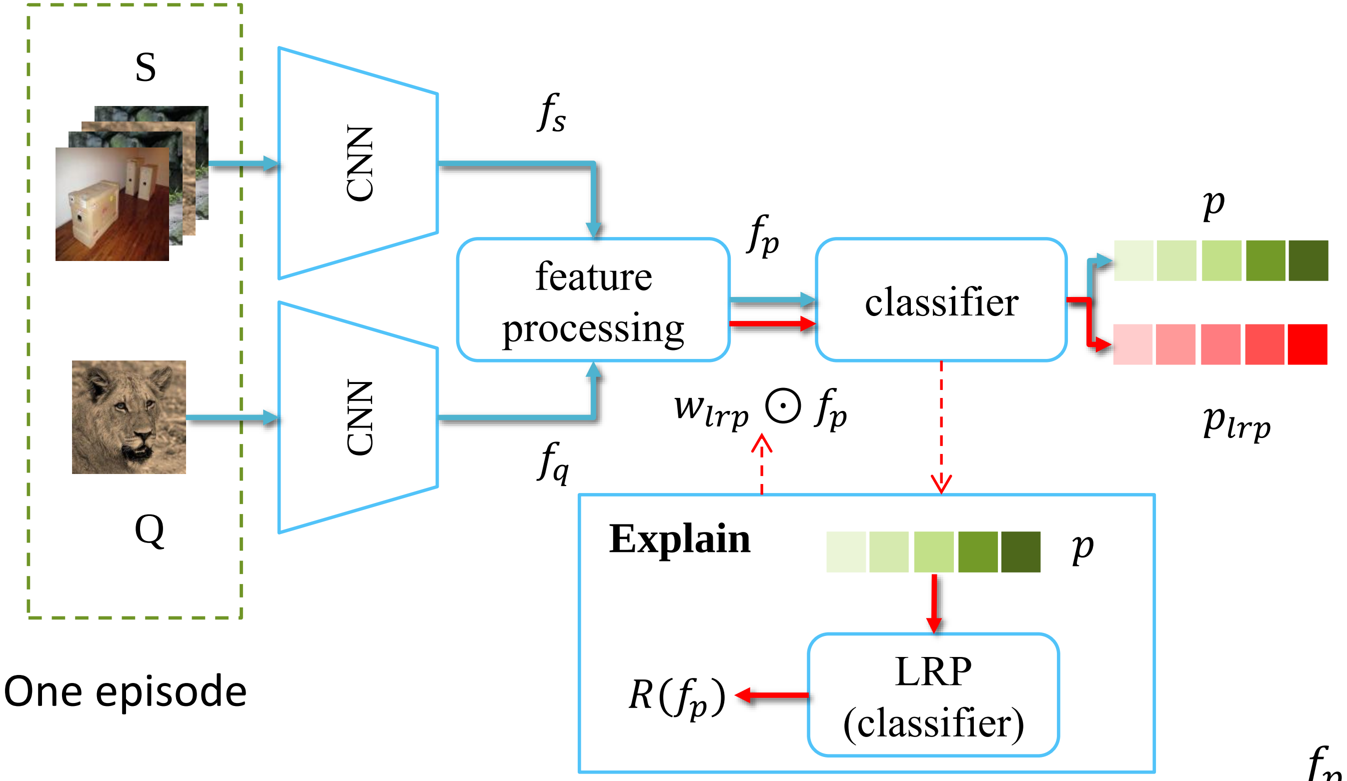
label



Outlines

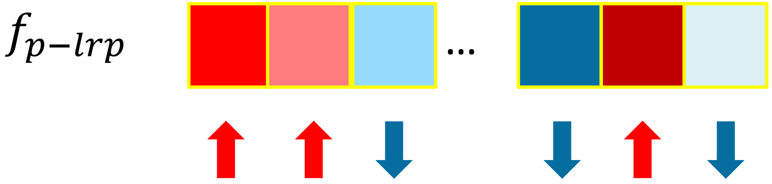
- Few-shot classification and the challenges of ***cross-domain*** few-shot classification.
- **Interpreting** few-shot classification models with **LRP**.
- **Explanation-guided training** for metric-based few-shot classification
- Performance and effects
- Conclusion

Explanation-Guided Training



$$L = \xi L_{ce}(y, p) + \lambda L_{ce}(y, p_{lrp})$$

$$w_{lrp} = 1 + R(f_p)$$
$$f_{p-lrp} = w_{lrp} \odot f_p$$

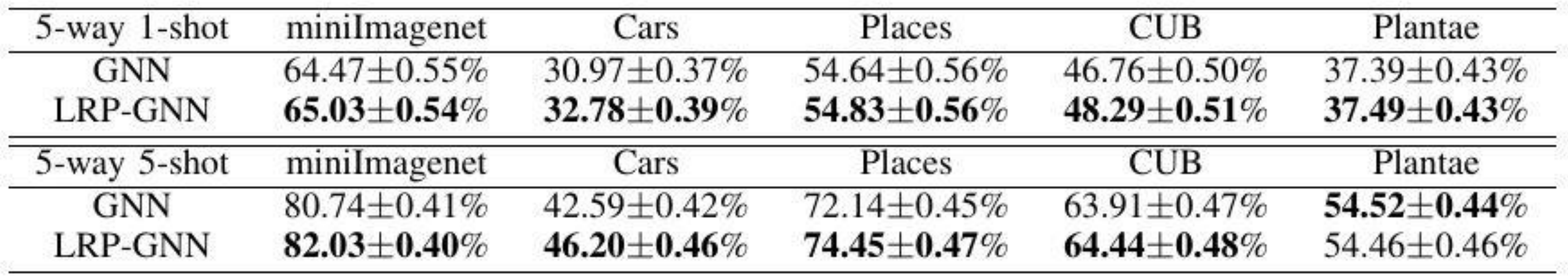


Outlines

- Few-shot classification and the challenges of ***cross-domain*** few-shot classification.
- **Interpreting** few-shot classification models with **LRP**.
- **Explanation-guided training** for metric-based few-shot classification
- Performance and effects
- Conclusion

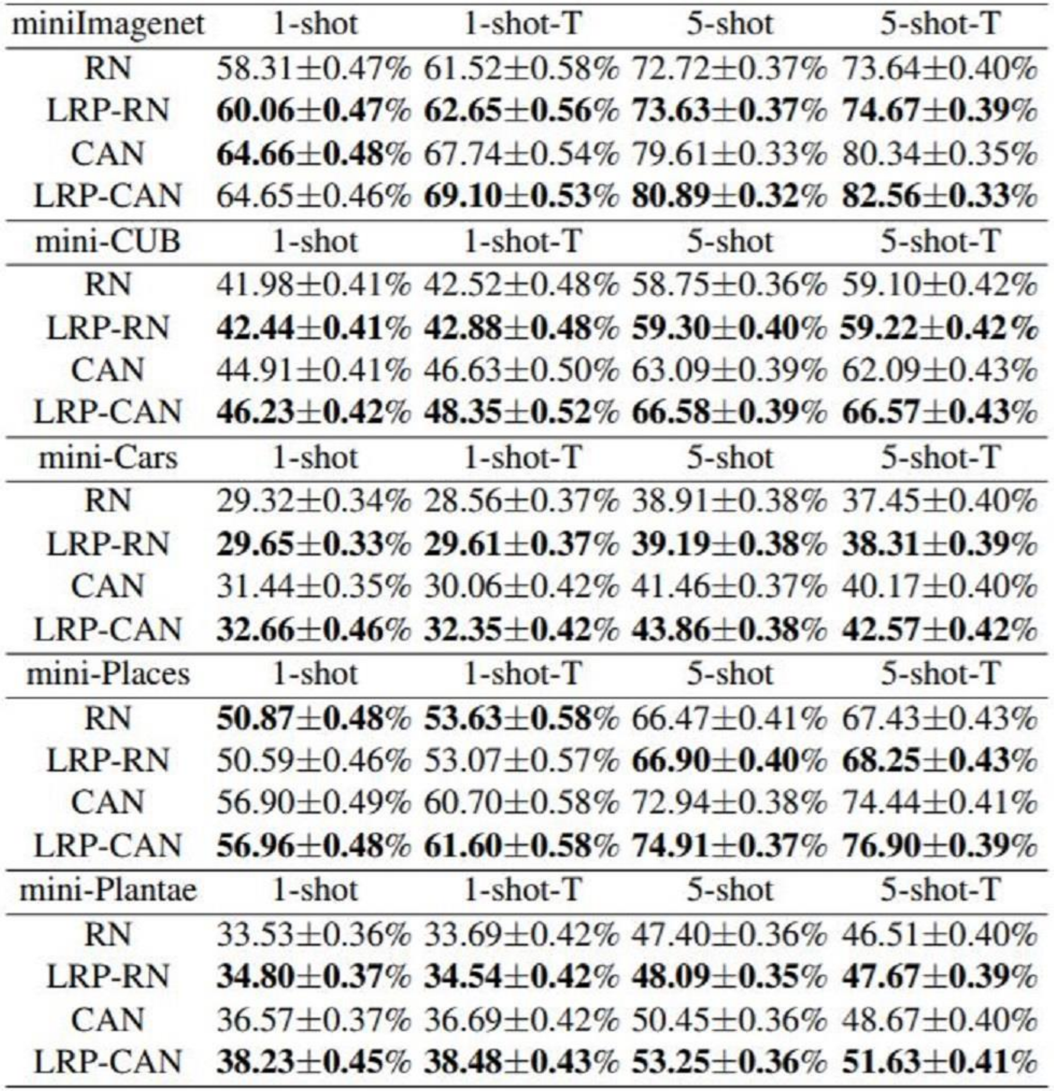
Performance and Effects

The performance of explanation-guided training on **GNN** on four cross domain datasets.



5-way 1-shot	miniImagenet	Cars	Places	CUB	Plantae
GNN	64.47±0.55%	30.97±0.37%	54.64±0.56%	46.76±0.50%	37.39±0.43%
LRP-GNN	**65.03±0.54%**	**32.78±0.39%**	**54.83±0.56%**	**48.29±0.51%**	**37.49±0.43%**
5-way 5-shot	miniImagenet	Cars	Places	CUB	Plantae
GNN	80.74±0.41%	42.59±0.42%	72.14±0.45%	63.91±0.47%	**54.52±0.44%**
LRP-GNN	**82.03±0.40%**	**46.20±0.46%**	**74.45±0.47%**	**64.44±0.48%**	54.46±0.46%

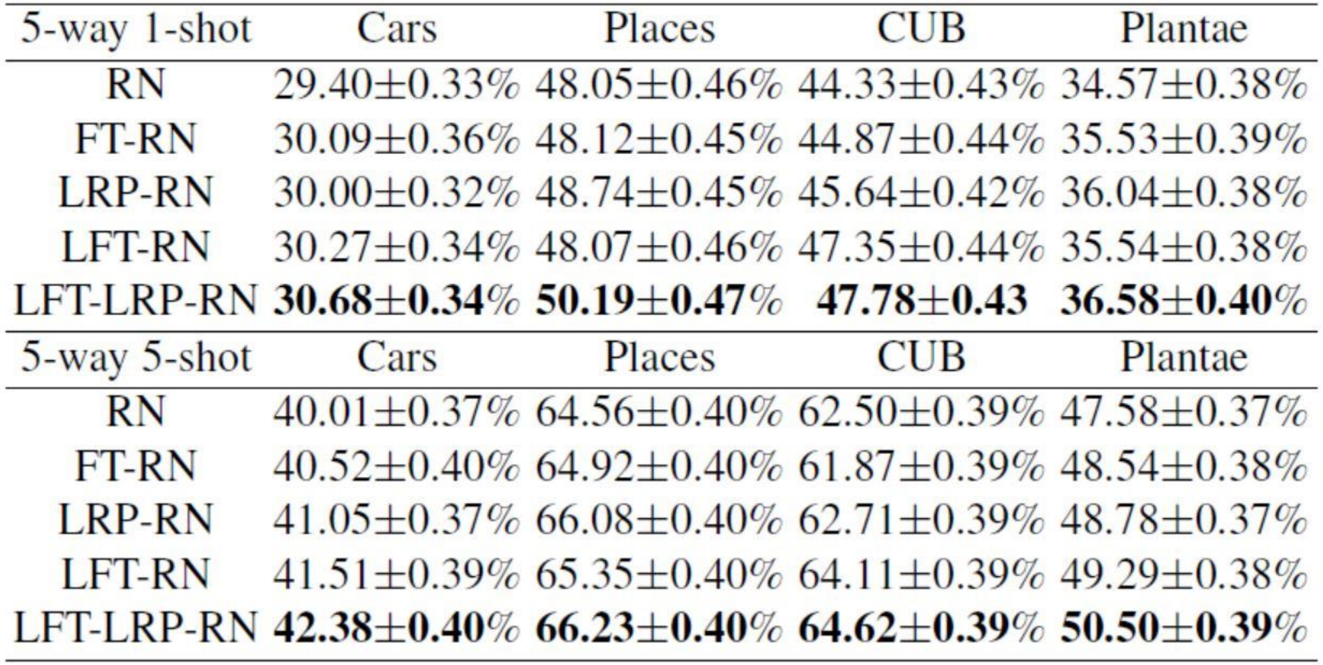
The performance of explanation-guided training on **RelationNet** (RN), **cross attention network** (CAN) on four cross domain datasets



miniImagenet	1-shot	1-shot-T	5-shot	5-shot-T
RN	58.31±0.47%	61.52±0.58%	72.72±0.37%	73.64±0.40%
LRP-RN	**60.06±0.47%**	**62.65±0.56%**	**73.63±0.37%**	**74.67±0.39%**
CAN	**64.66±0.48%**	67.74±0.54%	79.61±0.33%	80.34±0.35%
LRP-CAN	64.65±0.46%	**69.10±0.53%**	**80.89±0.32%**	**82.56±0.33%**
mini-CUB	1-shot	1-shot-T	5-shot	5-shot-T
RN	41.98±0.41%	42.52±0.48%	58.75±0.36%	59.10±0.42%
LRP-RN	**42.44±0.41%**	**42.88±0.48%**	**59.30±0.40%**	**59.22±0.42%**
CAN	44.91±0.41%	46.63±0.50%	63.09±0.39%	62.09±0.43%
LRP-CAN	**46.23±0.42%**	**48.35±0.52%**	**66.58±0.39%**	**66.57±0.43%**
mini-Cars	1-shot	1-shot-T	5-shot	5-shot-T
RN	29.32±0.34%	28.56±0.37%	38.91±0.38%	37.45±0.40%
LRP-RN	**29.65±0.33%**	**29.61±0.37%**	**39.19±0.38%**	**38.31±0.39%**
CAN	31.44±0.35%	30.06±0.42%	41.46±0.37%	40.17±0.40%
LRP-CAN	**32.66±0.46%**	**32.35±0.42%**	**43.86±0.38%**	**42.57±0.42%**
mini-Places	1-shot	1-shot-T	5-shot	5-shot-T
RN	**50.87±0.48%**	**53.63±0.58%**	66.47±0.41%	67.43±0.43%
LRP-RN	50.59±0.46%	53.07±0.57%	**66.90±0.40%**	**68.25±0.43%**
CAN	56.90±0.49%	60.70±0.58%	72.94±0.38%	74.44±0.41%
LRP-CAN	**56.96±0.48%**	**61.60±0.58%**	**74.91±0.37%**	**76.90±0.39%**
mini-Plantae	1-shot	1-shot-T	5-shot	5-shot-T
RN	33.53±0.36%	33.69±0.42%	47.40±0.36%	46.51±0.40%
LRP-RN	**34.80±0.37%**	**34.54±0.42%**	**48.09±0.35%**	**47.67±0.39%**
CAN	36.57±0.37%	36.69±0.42%	50.45±0.36%	48.67±0.40%
LRP-CAN	**38.23±0.45%**	**38.48±0.43%**	**53.25±0.36%**	**51.63±0.41%**

Combining with Other Methods

The combination of explanation-guided training and
learned feature-wise transformation (LFT)



5-way 1-shot	Cars	Places	CUB	Plantae
RN	29.40±0.33%	48.05±0.46%	44.33±0.43%	34.57±0.38%
FT-RN	30.09±0.36%	48.12±0.45%	44.87±0.44%	35.53±0.39%
LRP-RN	30.00±0.32%	48.74±0.45%	45.64±0.42%	36.04±0.38%
LFT-RN	30.27±0.34%	48.07±0.46%	47.35±0.44%	35.54±0.38%
LFT-LRP-RN	**30.68±0.34%**	**50.19±0.47%**	**47.78±0.43**	**36.58±0.40%**
5-way 5-shot	**Cars**	**Places**	**CUB**	**Plantae**
RN	40.01±0.37%	64.56±0.40%	62.50±0.39%	47.58±0.37%
FT-RN	40.52±0.40%	64.92±0.40%	61.87±0.39%	48.54±0.38%
LRP-RN	41.05±0.37%	66.08±0.40%	62.71±0.39%	48.78±0.37%
LFT-RN	41.51±0.39%	65.35±0.40%	64.11±0.39%	49.29±0.38%
LFT-LRP-RN	**42.38±0.40%**	**66.23±0.40%**	**64.62±0.39%**	**50.50±0.39%**

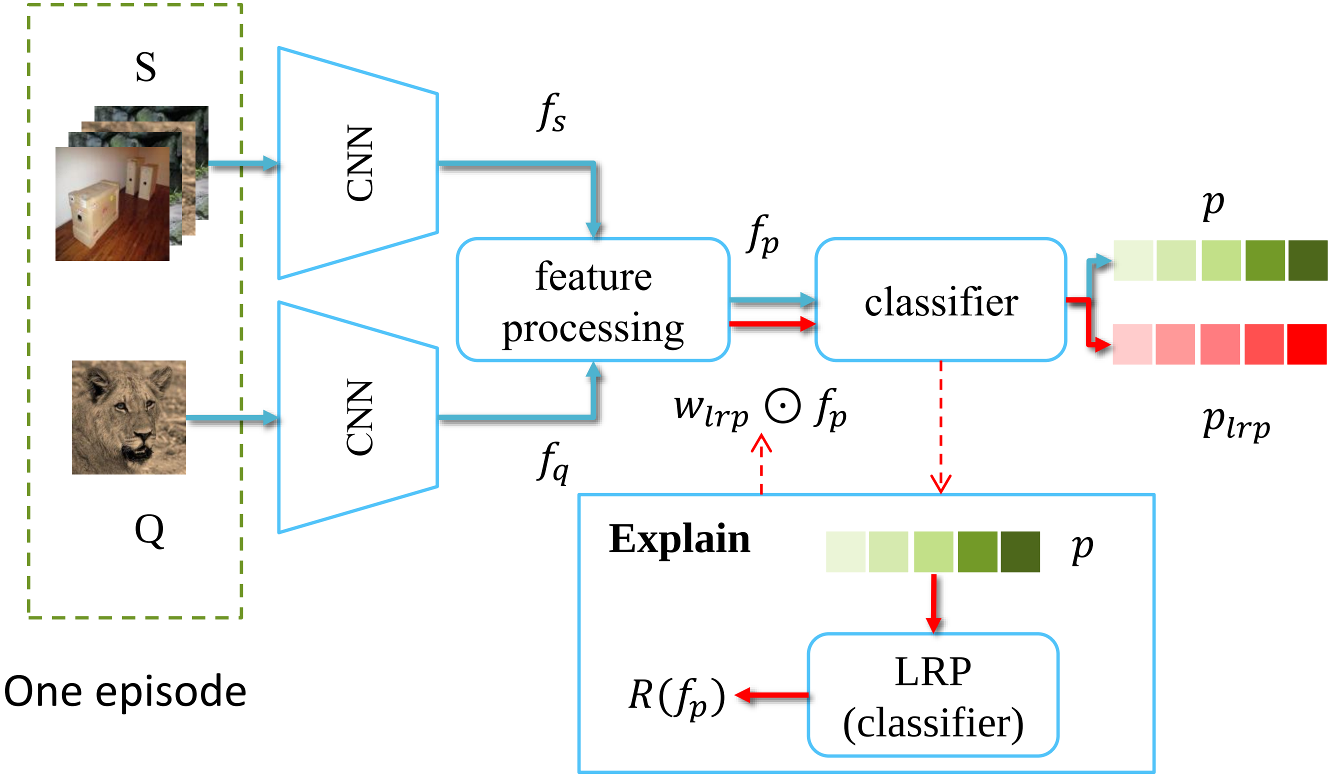
Outlines

- Few-shot classification and the challenges of ***cross-domain*** few-shot classification.
- **Interpreting** few-shot classification models with **LRP**.
- **Explanation-guided training** for metric-based few-shot classification
- Performance and effects
- Conclusion

Conclusion

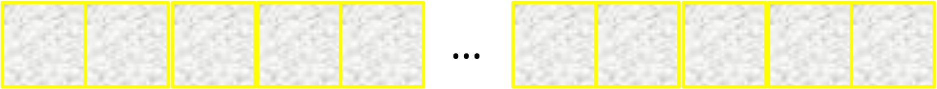
- We **Interpret** few-shot classification models with **LRP**.
- We propose **Explanation-guided training** for metric-based few-shot classification
- **Explanation-guided training** improves the performance on cross-domain few-shot classification tasks.
- **Explanation-guided training** can be combined with other methods such as **LFT**.

Performance and Effects



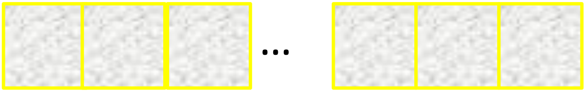
One episode

f_p

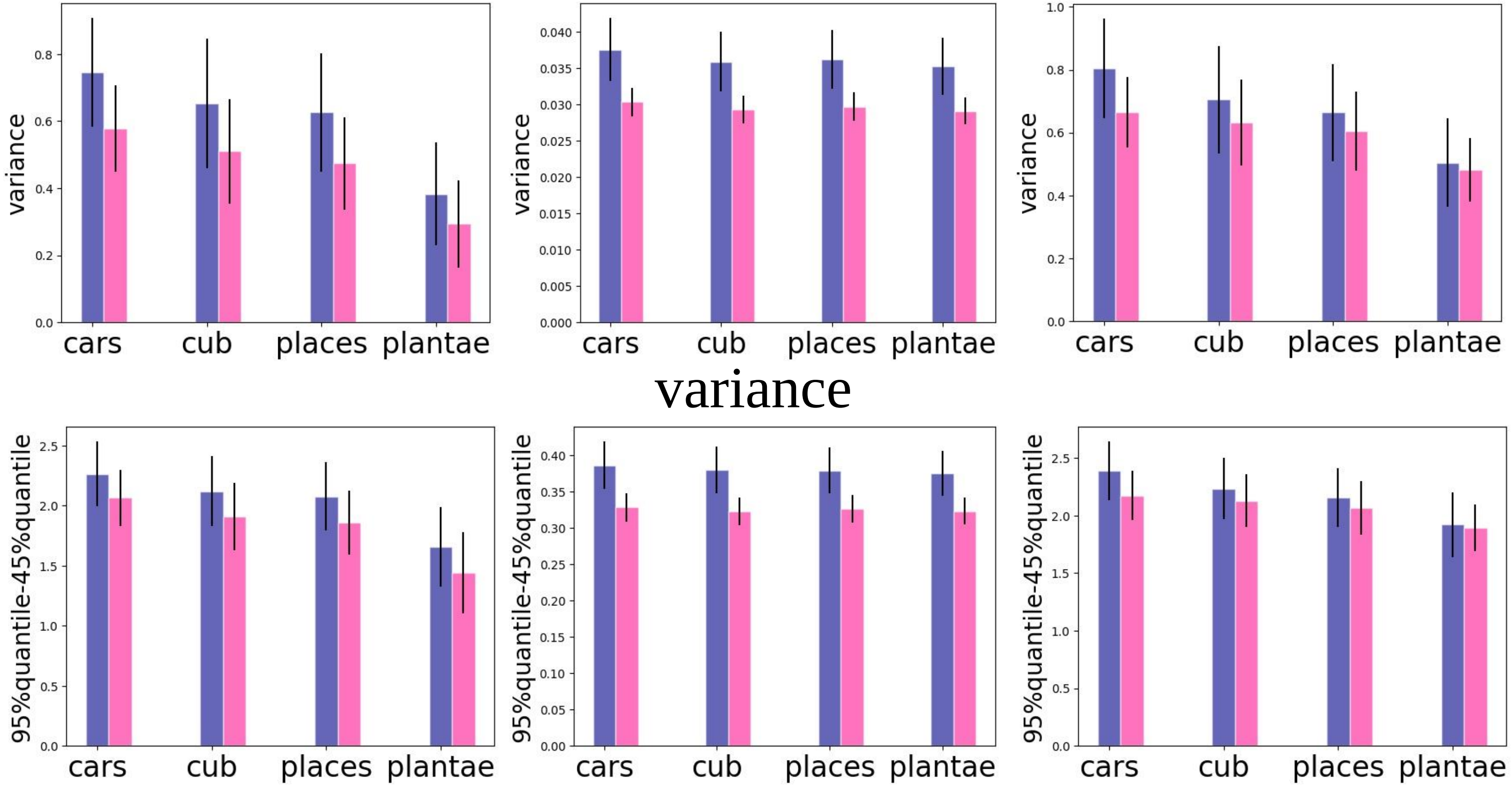


Performance and Effects

f_p



GNN CAN RelationNet



variance

difference between 95%quantile and 45%quantile

■ standard ■ explanation-guided