NEURAL INFORMATION PROCESSING SYSTEMS

DUAL

Dual Alignment Framework for Few-Shot Learning with Inter-Set and Intra-Set Shifts

Siyang Jiang, Rui Fang, Hsi-Wen Chen, Wei Ding, Guoliang Xing, Ming-Syan Chen

The Chinese University of Hong Kong
National Taiwan University

香港中文大學
The Chinese University of Hong Kong

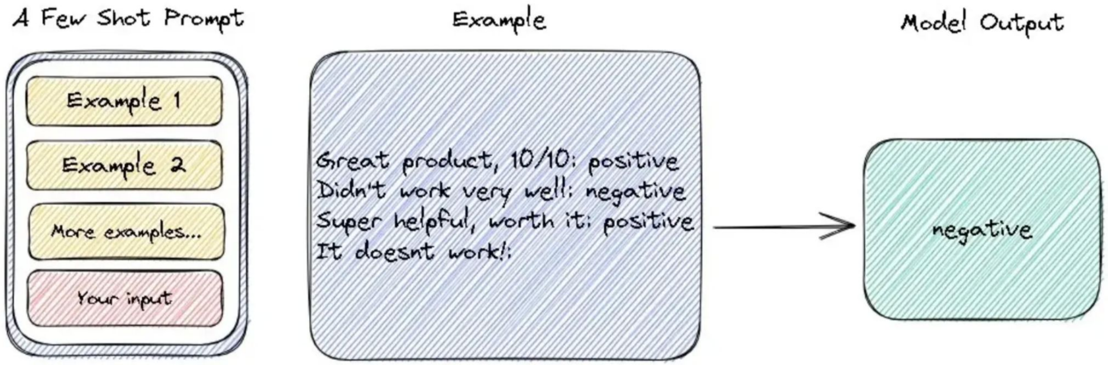
國立臺灣大學
National Taiwan University

Outline

- Background and Preliminary
- Motivation and Method
- Experiments
- Conclusion
- Acknowledgements

Background

Why Few-shot Learning (FSL)

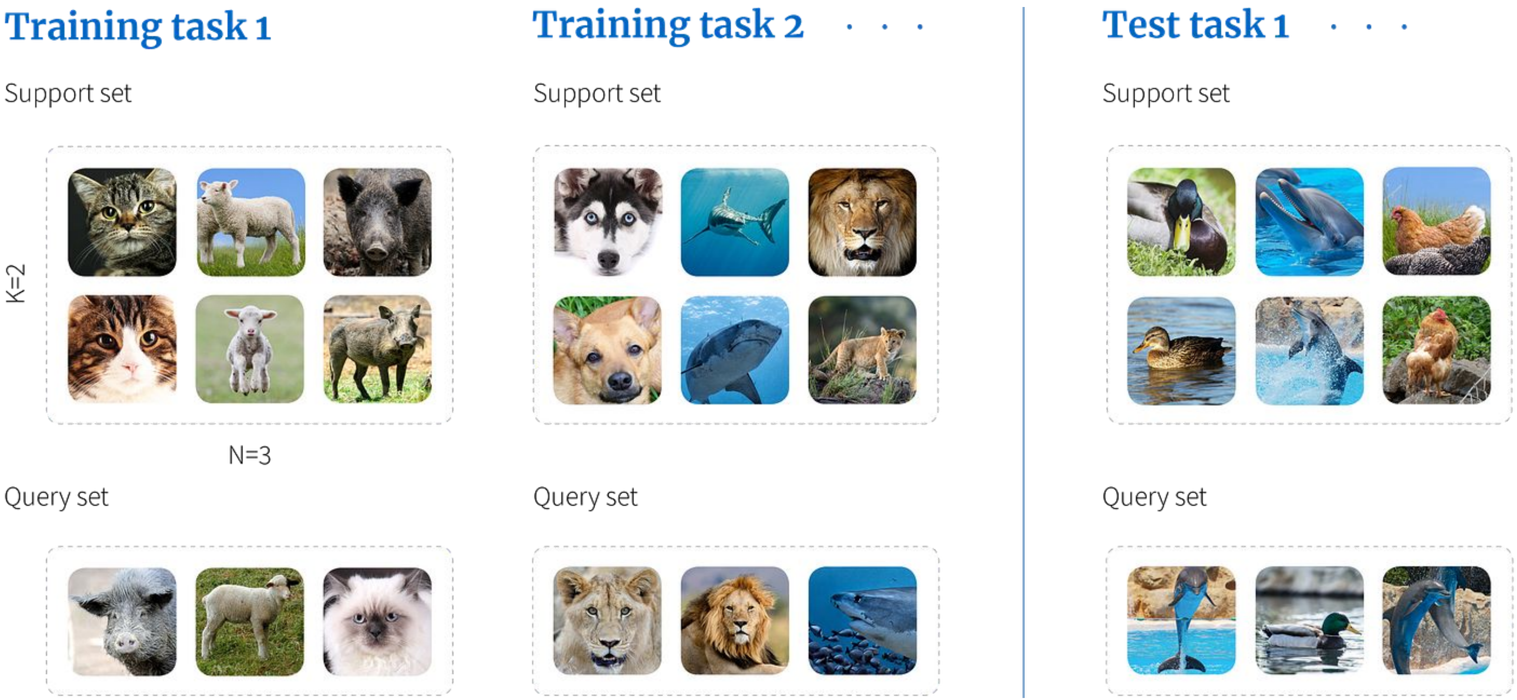


- Real-world tasks often have scarce labels and open-set classes.
- In FSL, it remains accurate under unpredictable shifts with minimal labels and compute.

Source From https://medium.com/ubiai-nlp/step-by-step-guide-to-mastering-few-shot-learning-a673054167a0

Preliminary

N-way K-shot



Source From https://medium.com/ai%C2%B3-theory-practice-business/what-is-few-shot-learning-4b2842646b47

Challenges

Dual Support Query Shift (DSQS)

Challenges

Dual Support Query Shift (DSQS)

DSQS denotes that it has **inter-set shifts and intra-set shifts** in Support set and Query set.

Challenges

Dual Support Query Shift (DSQS)

DSQS denotes that it has **inter-set shifts and intra-set shifts** in Support set and Query set.

Inter-set shift: support and query come from different domains (device, lighting, weather).

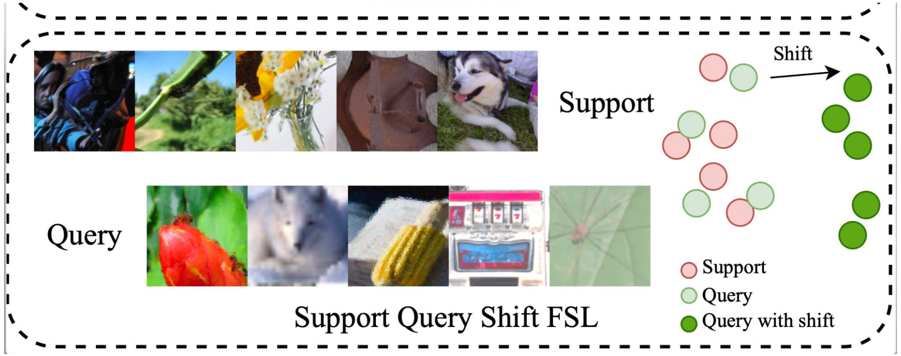
Intra-set shift: instances within the same support or query set suffer distinct disturbances.

Challenges

Dual Support Query Shift (DSQS)

DSQS denotes that it has
inter-set shifts and intra-set shifts

Obs. 1: We notice that two shifts catastrophically drop the model performance.

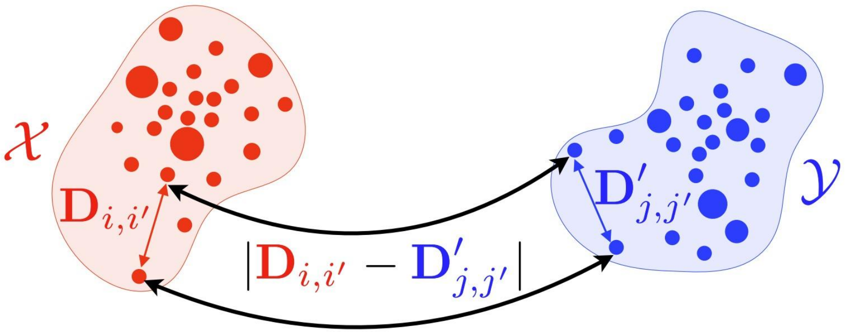


Inter-set shift: support and query come from different domains (device, lighting, weather).

Intra-set shift: instances within the same support or query set suffer distinct disturbances.

Motivation

Optimal Transport (OT)



Remarks: OT seeks the most *cost-efficient* way to move one distribution of mass (probability) to another while exactly conserving total mass.

Source From https://moscot.readthedocs.io/en/latest/ot_background.html

Motivation

Theoretical Insight

Proposition 1 (OT cost under first-order Gaussian approximation). *Approximating each domain by its first-order moments gives*

$$W_2^2(\mathcal{S}, \mathcal{Q}) \;=\; \underbrace{\|\mu_{\mathrm{s}} - \mu_{\mathrm{q}}\|_2^2}_{\textit{inter-set mean gap}} \;+\; \underbrace{\mathrm{tr}\big(\Sigma_{\mathrm{s}} + \Sigma_{\mathrm{q}} - 2(\Sigma_{\mathrm{s}}^{1/2}\Sigma_{\mathrm{q}}\Sigma_{\mathrm{s}}^{1/2})^{1/2}\big)}_{\textit{inter-set covariance gap}}\,. \tag{3}$$

Thus, the transport cost grows monotonically with (i) the interset mean gap $\|\mu_{\mathrm{s}} - \mu_{\mathrm{q}}\|_2$ and (ii) the intraset spreads $\mathrm{tr}(\Sigma_{\mathrm{s}})$ and $\mathrm{tr}(\Sigma_{\mathrm{q}})$.[3]

Proposition 2 (Error of transported embeddings). *Let $\hat{\phi}(x_{q,i})$ be the transported query embedding obtained from the clean OT plan in Eq. (1), and $\hat{\phi}_\sigma(x_{q,i})$ its noisy counterpart. Assume additive Gaussian noise $\eta \sim \mathcal{N}(0, \sigma_\diamond^2 I)$ is independently injected in both support and query domains $\diamond \in \{\mathrm{s}, \mathrm{q}\}$. Then,*

$$\mathbb{E}[\|\hat{\phi}(x_{q,i}) - \hat{\phi}_\sigma(x_{q,i})\|_2^2] = d(\sigma_{\mathrm{s}}^2 + \sigma_{\mathrm{q}}^2).$$

Larger noise levels $\sigma_{\mathrm{s}}, \sigma_{\mathrm{q}}$ therefore *magnify* the risk of a mismatched OT plan, ultimately degrading classification accuracy.

Motivation

Theoretical Insight

Proposition 1 (OT cost under first-order Gaussian approximation). *Approximating each domain by its first-order moments gives*

$$W_2^2(\mathcal{S}, \mathcal{Q}) = \underbrace{\|\mu_\mathrm{s} - \mu_\mathrm{q}\|_2^2}_{\textit{inter-set mean gap}} + \underbrace{\mathrm{tr}\big(\Sigma_\mathrm{s} + \Sigma_\mathrm{q} - 2(\Sigma_\mathrm{s}^{1/2}\Sigma_\mathrm{q}\Sigma_\mathrm{s}^{1/2})^{1/2}\big)}_{\textit{inter-set covariance gap}}. \tag{3}$$

Obs. 2: Domain misalignment and feature noise strongly affect the performance of OT.

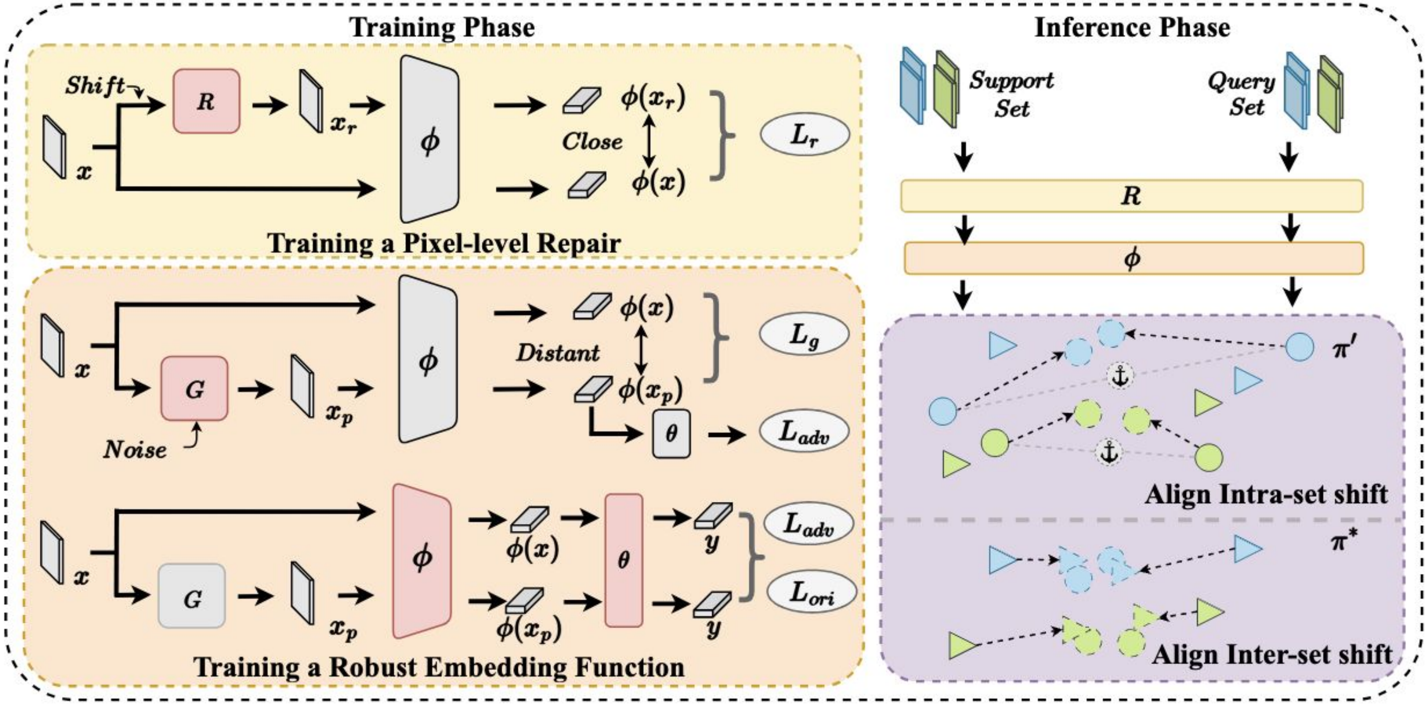
obtained from the clean OT plan in Eq. (1), and $\hat{\phi}_\sigma(x_{q,i})$ its noisy counterpart. Assume additive Gaussian noise $\eta \sim \mathcal{N}(0, \sigma_\diamond^2 I)$ is independently injected in both support and query domains $\diamond \in \{\mathrm{s}, \mathrm{q}\}$. Then,

$$\mathbb{E}[\|\hat{\phi}(x_{q,i}) - \hat{\phi}_\sigma(x_{q,i})\|_2^2] = d(\sigma_\mathrm{s}^2 + \sigma_\mathrm{q}^2).$$

Larger noise levels $\sigma_\mathrm{s}, \sigma_\mathrm{q}$ therefore *magnify* the risk of a mismatched OT plan, ultimately degrading classification accuracy.

Methodology

Overview of DUAL



Methodology

DUAL Adversarial Training (Dual AT)

- The goal of Dual AT is to *Make Features Clean and Robust*

Repairer R (pixel level Repairer) : denoise shifted images; train by minimizing $M\big(\phi(x_r), \phi(x)\big)$ to preserve semantics.

Generator G (semantic hard examples): maximize difficulty $M\big(\phi(G(x)), \phi(x)\big)$, keeping label via KL consistency.

Embedding ϕ : trained with $\mathcal{L}_{\text{ori}} = \mathrm{KL}\big(\theta(\phi(x)), y\big)$ and $\mathcal{L}_{\text{adv}} = \mathrm{KL}\big(\theta(\phi(G(x))), y\big)$

Methodology

DUAL Optimal Transportation (Dual OT)

- The goal of Dual OT is to *Align Within and Across Sets*

Stage-1 (π', intra-set) : transport support instances to class-wise centroids; adopting negative-entropy regularization.

Stage-2 (π^*, inter-set) : align queries to transported support S' (anchors) using the same regularized OT.

Experiments

Setup

- 3 Datasets
- 10 Baselines
- 1-shot and 5 shots settings

Exp-1

Main Results

Methods	CIFAR-100	mini-ImageNet	Tiered-ImageNet	CIFAR-100	mini-ImageNet	Tiered-ImageNet
	1-shot			5-shot		
MatchingNet [45]	$30.26_{\pm 0.38}$	$43.62_{\pm 0.47}$	$30.01_{\pm 0.41}$	$40.35_{\pm 0.33}$	$56.24_{\pm 0.37}$	$35.05_{\pm 0.36}$
ProtoNet [42]	$28.53_{\pm 0.30}$	$43.84_{\pm 0.44}$	$30.15_{\pm 0.41}$	$41.59_{\pm 0.41}$	$59.83_{\pm 0.42}$	$43.41_{\pm 0.43}$
TransPropNet [31]	$31.01_{\pm 0.34}$	$24.22_{\pm 0.29}$	$24.18_{\pm 0.32}$	$37.06_{\pm 0.40}$	$25.93_{\pm 0.29}$	$35.48_{\pm 0.37}$
FTNet [9]	$22.36_{\pm 0.21}$	$37.04_{\pm 0.44}$	$22.01_{\pm 0.30}$	$26.19_{\pm 0.25}$	$49.14_{\pm 0.36}$	$24.50_{\pm 0.23}$
AQ [12]	$35.86_{\pm 0.54}$	$31.59_{\pm 0.44}$	$30.24_{\pm 0.40}$	$53.93_{\pm 0.55}$	$43.85_{\pm 0.49}$	$38.54_{\pm 0.42}$
TP [3]	$30.89_{\pm 0.42}$	$45.66_{\pm 0.55}$	$29.34_{\pm 0.43}$	$45.50_{\pm 0.37}$	$62.32_{\pm 0.38}$	$41.92_{\pm 0.39}$
PGADA [21]	$34.90_{\pm 0.45}$	$50.37_{\pm 0.57}$	$28.47_{\pm 0.40}$	$49.45_{\pm 0.38}$	$61.09_{\pm 0.39}$	$40.73_{\pm 0.34}$
AQP [1]	$31.68_{\pm 0.39}$	$30.59_{\pm 0.43}$	$30.40_{\pm 0.40}$	$45.09_{\pm 0.46}$	$42.65_{\pm 0.57}$	$45.34_{\pm 0.60}$
RAS [35]	$36.98_{\pm 0.38}$	$50.40_{\pm 0.32}$	$31.05_{\pm 0.40}$	$50.02_{\pm 0.21}$	$63.95_{\pm 0.40}$	$43.98_{\pm 0.42}$
SSL-ProtoNet [28]	$36.00_{\pm 0.38}$	$28.59_{\pm 0.30}$	$29.31_{\pm 0.48}$	$48.74_{\pm 0.37}$	$36.56_{\pm 0.32}$	$35.65_{\pm 0.37}$
DUAL-P	$38.93_{\pm 0.50}$	$53.00_{\pm 0.60}$	$34.29_{\pm 0.50}$	$54.47_{\pm 0.40}$	$67.83_{\pm 0.40}$	$47.81_{\pm 0.41}$
DUAL-M	$39.35_{\pm 0.51}$	$54.44_{\pm 0.59}$	$35.29_{\pm 0.37}$	$50.11_{\pm 0.40}$	$64.04_{\pm 0.42}$	$42.96_{\pm 0.38}$

Exp-1

- **Consistent gains; up to relatively 25.66% average improvement across three datasets.**
- **Surpasses relatively robust FSL (SSL-ProtoNet) by ~35.65% on average.**
- **Strong on 1-shot and 5-shot;**

Exp-2

Ablation Study

Techniques Variants	CIFAR-100	mini-ImageNet	Tiered-ImageNet	CIFAR-100	mini-ImageNet	Tiered-ImageNet
	1-shot			5-shot		
w./o. dual AT & OT	$27.43_{\pm 0.32}$	$43.93_{\pm 0.47}$	$27.85_{\pm 0.35}$	$41.97_{\pm 0.41}$	$63.60_{\pm 0.45}$	$40.48_{\pm 0.40}$
w./o. dual AT	$31.36_{\pm 0.41}$	$53.43_{\pm 0.59}$	$30.76_{\pm 0.43}$	$42.00_{\pm 0.44}$	$66.22_{\pm 0.46}$	$40.84_{\pm 0.40}$
w./o. dual OT	$34.63_{\pm 0.40}$	$40.88_{\pm 0.45}$	$27.54_{\pm 0.36}$	$53.20_{\pm 0.44}$	$66.69_{\pm 0.43}$	$30.57_{\pm 0.34}$
w./o. G	$35.98_{\pm 0.28}$	$43.74_{\pm 0.79}$	$29.32_{\pm 0.37}$	$47.10_{\pm 0.47}$	$61.22_{\pm 0.78}$	$43.95_{\pm 0.49}$
w./o. R	$27.47_{\pm 0.36}$	$44.12_{\pm 0.43}$	$26.73_{\pm 0.26}$	$35.05_{\pm 0.39}$	$62.33_{\pm 0.38}$	$37.92_{\pm 0.32}$
Fixed G	$38.48_{\pm 0.50}$	$55.35_{\pm 0.61}$	$31.12_{\pm 0.47}$	$52.47_{\pm 0.47}$	$66.91_{\pm 0.47}$	$42.54_{\pm 0.40}$
Enc shift to ϕ	$34.56_{\pm 0.38}$	$49.37_{\pm 0.50}$	$24.26_{\pm 0.26}$	$45.98_{\pm 0.38}$	$62.55_{\pm 0.39}$	$29.11_{\pm 0.29}$
TP + R	$32.03_{\pm 0.36}$	$48.58_{\pm 0.53}$	$28.52_{\pm 0.39}$	$46.13_{\pm 0.40}$	$64.25_{\pm 0.40}$	$41.22_{\pm 0.38}$
DUAL-P	$38.93_{\pm 0.50}$	$53.00_{\pm 0.59}$	$34.29_{\pm 0.50}$	$54.47_{\pm 0.40}$	$67.83_{\pm 0.40}$	$47.81_{\pm 0.41}$

Exp-2

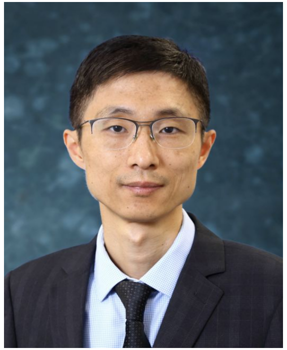
- **Remove Dual AT or Dual OT → large drops;**
  - **without R: −22.54%;**
  - **without G: −11.80%;**
  - **fixing G: −3.20%.**
- **Dual AT alone adds ~11.59%**
  - **(largest gain on Tiered-ImageNet 1-shot).**

Conclusion

- Propose the Dual Support-Query Shift (DSQS) challenge, which investigates inter-set and intra-set shift problem in FSL.
- Theoretically prove that both shifts can misguide the domain alignment process by optimal transportation.
- Propose the DUal ALignment Framework (DUAL), which leverages a repairer and a robust embedding function adversarially trained by a generator to obtain clean features.
- Extensive experiments demonstrate that DUAL outperforms 10 state-of-the-art methods, achieving an average improvement of 25.66% across three datasets.

Acknowledgement

- My Ph.D. supervisor is Prof. Guoliang Xing,
  and my master's supervisor is Prof. Ming-Syan Chen.



- Co-authors from AIoT Lab and NetDB Lab
  - Rui Fang, Hsi-Wen Chen, Wei Ding

NEURAL INFORMATION
PROCESSING SYSTEMS

Thanks for your listening

Siyang Jiang

E-mail: syjiang [AT] ie.cuhk.edu.hk

香港中文大學
The Chinese University of Hong Kong

國立臺灣大學
National Taiwan University