Context-Conditional Adaptation for Recognizing Unseen Classes in Unseen Domains

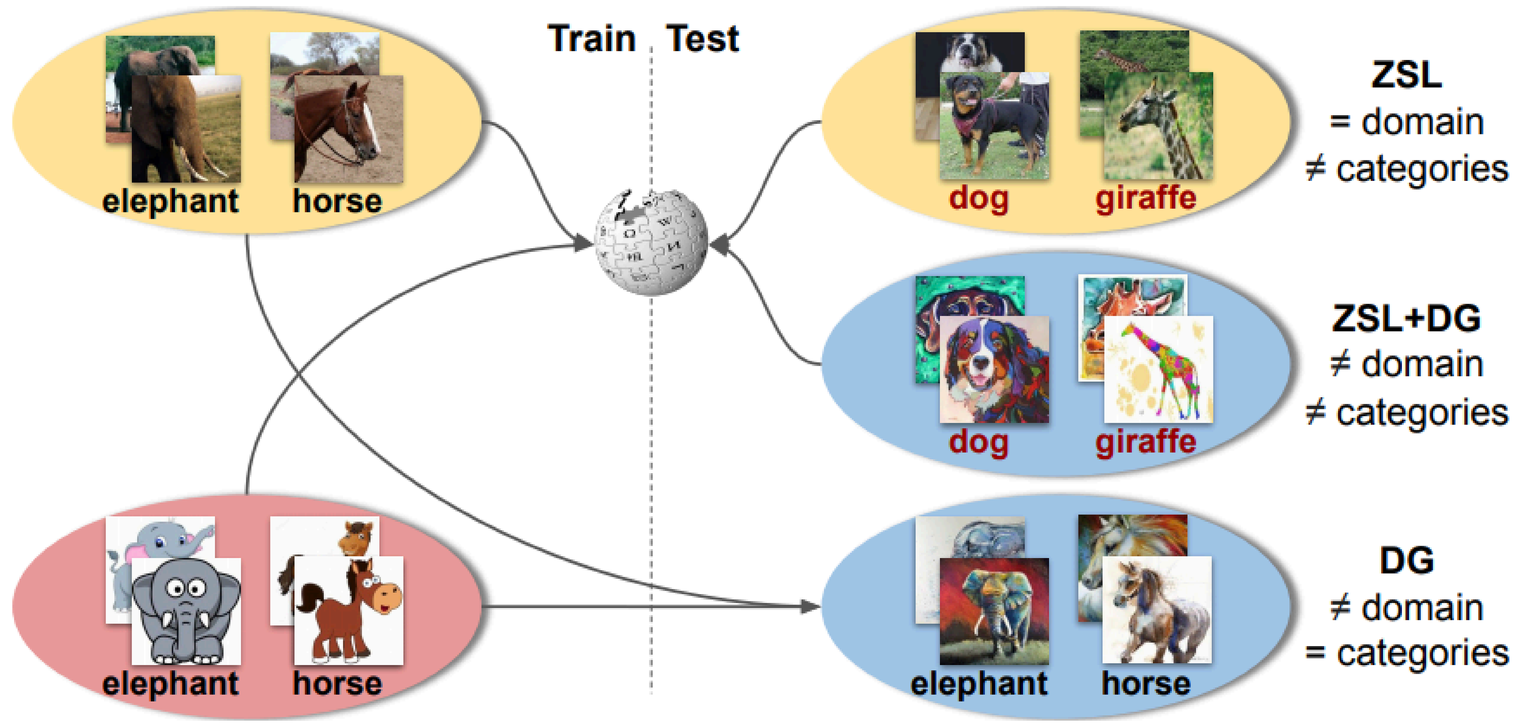
Puneet Mangla [1] **Shivam Chandhok** [2] Vineeth N Balasubramanian [1] Fahad Shahbaz Khan [2]

[1]Department of Computer Science, Indian Institute of Technology, Hyderabad [2]Mohamed bin Zayed University of Artificial Intelligence

ArXiv

Problem Definition and Contribution

Zero-Shot Domain Generalization: Tackling domain shift and semantic shift together. In ZSLDG, the model has access to a set of seen classes from multiple source domains and has to generalize to unseen classes in unseen target domains at test time.



Credits: CuMix [5]

Key Contributions:

- A generative framework that uses COntext COnditional Adaptive (COCOA) Batch Normalization to seamlessly integrate semantic and domain information.
- Demonstrate state-of-the-art performance for the ZSLDG setting.
- Illustrate COCOA's ability to encode both semantic and domain-specific information to generate features.

Notations and Preliminaries

Goal : train a classifier $\mathcal{C}$ which can recognize unseen classes in unseen domains at test-time.

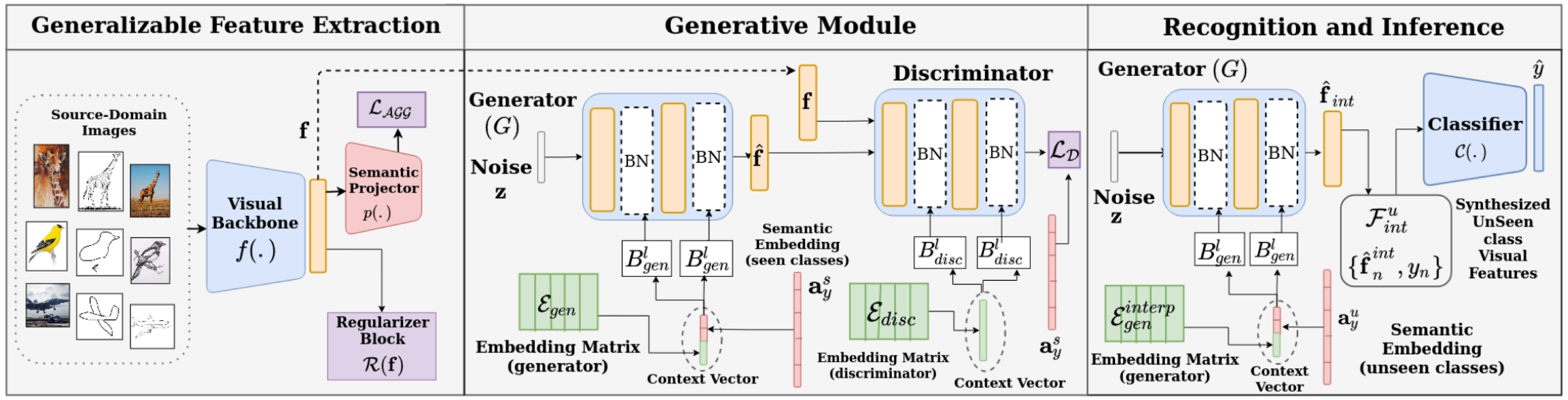
- Let $S^{Tr} = \{(\mathbf{x}, y, \mathbf{a}_y^s, d) \mid \mathbf{x} \in \mathcal{X}, y \in \mathcal{Y}^s, \mathbf{a}_y^s \in \mathcal{A}, d \in \mathcal{D}^s\}$ denote the training set.
- $S^{Ts} = \{(\mathbf{x}, y, \mathbf{a}_y^u, d) \mid \mathbf{x} \in \mathcal{X}, y \in \mathcal{Y}^u, \mathbf{a}_y^u \in \mathcal{A}, d \in \mathcal{D}^u\}$ denote the test set.
- The challenging ZSLDG setting assumes $\mathcal{Y}^s \cap \mathcal{Y}^u \equiv \emptyset$ and $\mathcal{D}^s \cap \mathcal{D}^u \equiv \emptyset$.

Motivation

Domain Shift : Feature representations that encode both domain-invariant and domain-specific information achieve improved recognition performance and generalization to new domains [1, 2].
Semantic Shift : Generative approaches [3, 4] outperform other methods (such as embedding-based methods) enabling better generalization to unseen classes at test time.

Motivated from these observations, **we aim to learn a generative model which can encode both class-level semantic and domain-specific information in the generated features**.

Our Approach: COntext-COnditional Adaptation (COCOA)



Step 1: *Generalizable Feature Extraction*: trains a visual backbone $f(.)$ to extract features that encodes domain-invariant and domain-specific info.

Step 2: *Generative Model*: learns a generative model which uses COCOA batch-normalization layers to fuse and integrate semantic and domain-specific info into generated features.

Context-Conditional Adaptive Batch-Normalization: We use a context-vector $\mathbf{c}$ to estimate batch-norm parameters as follows:

$$\gamma_{gen}^l, \beta_{gen}^l \leftarrow B_{gen}^l(\mathbf{c}) \;,\; \mathbf{f}_{gen}^{l+1} \leftarrow \gamma_{gen}^l \cdot \frac{\mathbf{f}_{gen}^l - \mu_{gen}^l}{\sqrt{(\sigma_{gen}^l)^2 + \epsilon}} + \beta_{gen}^l \text{where } \; \mathbf{c} = [\mathbf{a}_y^s, \mathbf{e}_d^{gen}]$$
$$\gamma_{disc}^l, \beta_{disc}^l \leftarrow B_{disc}^l(\mathbf{e}_d^{disc}) \;,\; \mathbf{f}_{disc}^{l+1} \leftarrow \gamma_{disc}^l \cdot \frac{\mathbf{f}_{disc}^l - \mu_{disc}^l}{\sqrt{(\sigma_{disc}^l)^2 + \epsilon}} + \beta_{disc}^l \qquad (1)$$

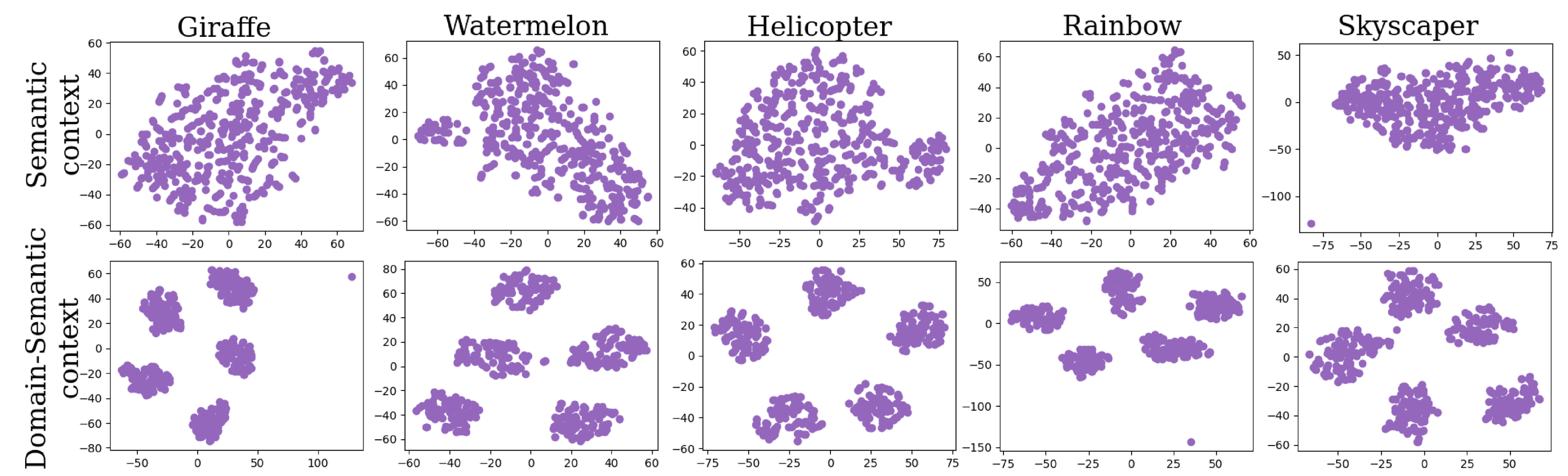
Step 3: *Inference*: generates visual features for unseen classes across domains (via domain interpolation) and trains a softmax classifier.

Experiments

Performance comparison with established baselines and state-of-art methods for ZSLDG problem setting on benchmark DomainNet dataset.

Method DG	Method ZSL	*clipart*	*infograph*	*painting*	*quickdraw*	*sketch*	Avg.
		Target Domain					
-	DEVISE	20.1	11.7	17.6	6.1	16.7	14.4
	ALE	22.7	12.7	20.2	6.8	18.5	16.2
	SPNet	26.0	16.9	23.8	8.2	21.8	19.4
DANN	DEVISE	20.5	10.4	16.4	7.1	15.1	13.9
	ALE	21.2	12.5	19.7	7.4	17.9	15.7
	SPNet	25.9	15.8	24.1	8.4	21.3	19.1
EpiFCR	DEVISE	21.6	13.9	19.3	7.3	17.2	15.9
	ALE	23.2	14.1	21.4	7.8	20.9	17.5
	SPNet	26.4	16.7	24.6	9.2	23.2	20.0
Mixup-img-only		25.2	16.3	24.4	8.7	21.7	19.2
Mixup-two-level		26.6	17	25.3	8.8	21.9	19.9
CuMix		<u>27.6</u>	17.8	25.5	9.9	22.6	20.7
f-clsWGAN		20.0	13.3	20.5	6.6	14.9	15.1
CuMix + f-clsWGAN		27.3	<u>17.9</u>	26.5	11.2	24.8	21.5
ROT + f-clsWGAN		27.5	17.4	26.4	11.4	24.6	21.4
COCOA_{AGG}		27.6	17.1	25.7	11.8	23.7	21.2
COCOA_{ROT}		**28.9**	**18.2**	<u>27.1</u>	**13.1**	**25.7**	**22.6**

Individual t-SNE visualization of synthesized image features by COCOA for randomly selected unseen classes (*Giraffe, Watermelon, Helicopter, Rainbow, Skyscraper*)



References

[1] Prithvijit Chattopadhyay, Y. Balaji, and Judy Hoffman.Learning to balance specificity and invariance for in and outof domain generalization.
[2] Seonguk Seo, Yumin Suh, D. Kim, Jongwoo Han, and B.Han. Learning to optimize domain specific normalizationfor domain generalization.
[3] Yuming Shen, J. Qin, and L. Huang. Invertible zero-shot recognition flows. In ECCV, 2020.
[4] Shivam Chandhok and V. Balasubramanian. Two-level adversarial visualsemantic coupling for generalized zero-shot learning. WACV, 2021.
[5] Massimiliano Mancini, Zeynep Akata, E. Ricci, and Barbara Caputo. Towards recognizing unseen categories in unseen domains. In ECCV, 2020.