Causality Compensated Attention for Contextual Biased Visual Recognition

Olaluwoye Olalekan

African Masters in Machine Intelligence (AMMI), Senegal

Authors: Ruyang Liu, Jingjia Huang, Thomas H. Li, Ge Li (2023)

29th, August, 2025

Overview

- Introduction
- Why Causality Matters in Image Recognition
- Causal View of Contextual Bias and Attention Mechanism.
- Causal Intervention: Theoretical Framework
- Methodology (IDA)
- Results and Discussion
- Conclusion

Introduction

- Attention mechanisms help models focus on important features in images.
- Key to improving accuracy in tasks like classification and object detection.
- **Problem:** Models often picks up context (background) instead of objects due to contextual bias.
- Leads to incorrect predictions when objects appear in unfamiliar contexts.
- The Authors propose a new attention module called IDA.

Aim and Objectives

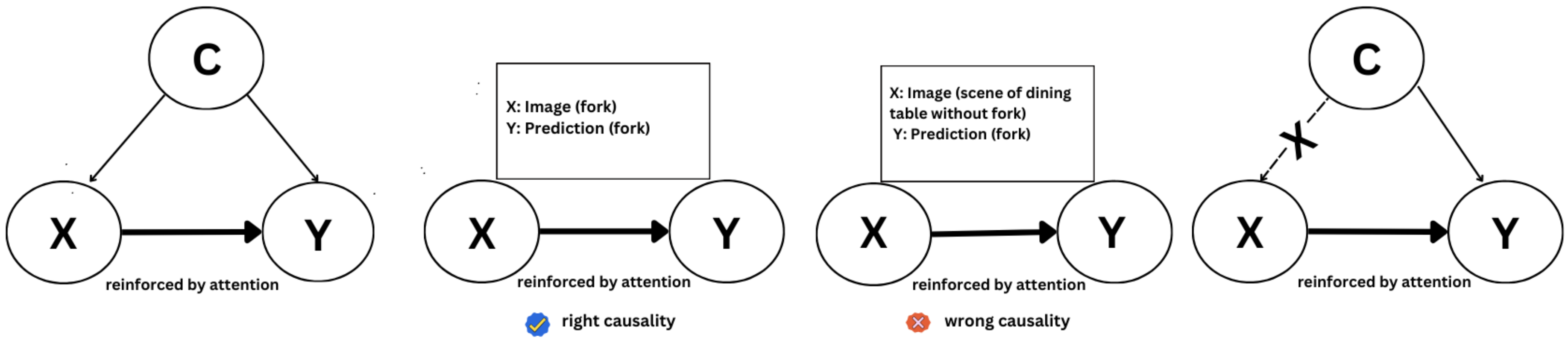
Aims: To improve attention mechanisms by reducing the impact of contextual bias.

The objectives are:

- Understand how current models are affected by contextual bias.
- Develop the IDA module to correct the bias.
- Evaluate the effectiveness of IDA on benchmark tasks (multi-label classification and object detection).

Causal View of Contextual Bias, Attention Mechanism and the Intervention

Figure 1: Demonstration of the causal view of contextual bias in visual recognition



(a) The confounding effect (X⟵C⟶Y)

(b) The role of attention mechanism

(c) Causal intervention

Causal Intervention: Theoretical Framework

- **Interventions** cut off misleading context, ensuring the model focuses on the relevant features (*Pearl, 2009*).
- Backdoor adjustment: Used to control for context that could falsely associate objects and predictions.
- The intervention equation can be expressed as:

$$P(Y|do(X)) = \sum_{c} P(Y|X, C=c) P(C=c)$$

Where: X is the object, Y is the prediction, and C represents the context.

- $do(X)$ refers to intervening directly on X, breaking the confounding effect of the context.
- $P(Y|X, C=c)$: The probability of Y given X and C.
- P(c) is the probability distribution of the confounder.

Solution: Intervention Dual Attention (IDA)

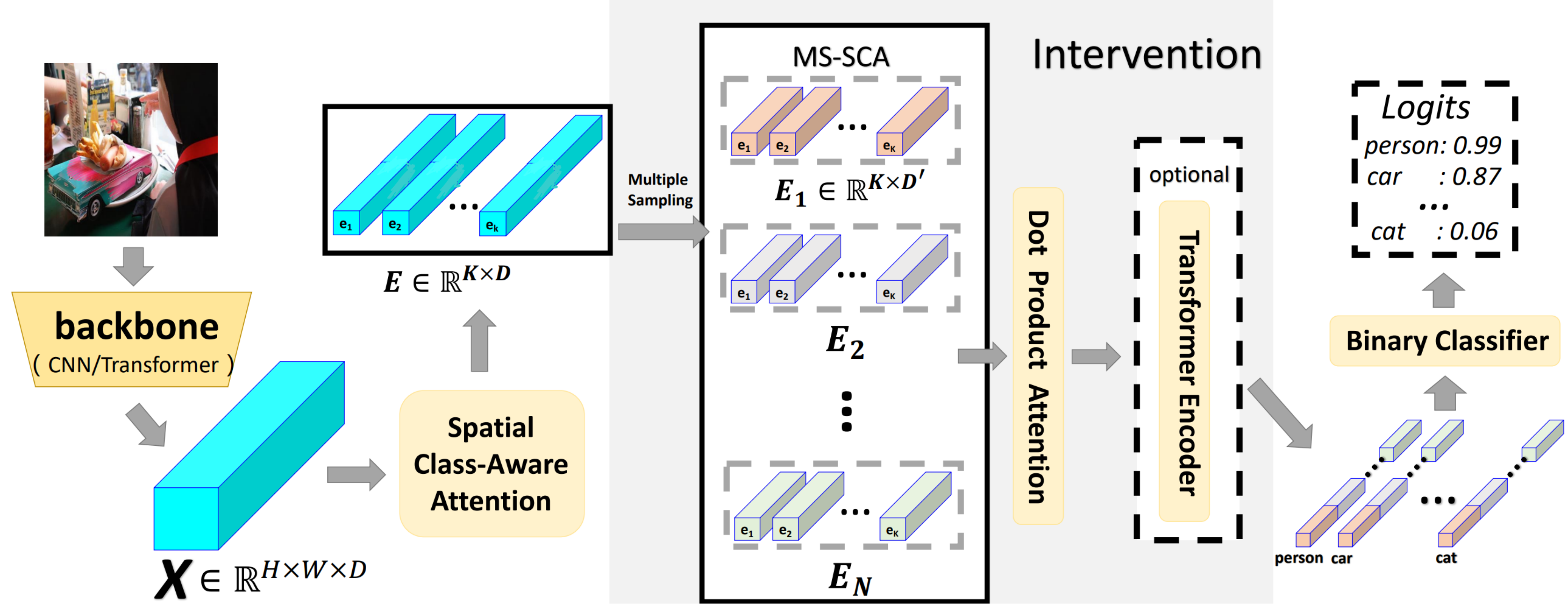


Fig. 2: Overview of the proposed model. X could be either image feature from visual backbone or ROI feature from detection backbone. The model is composed of the baseline attention (SCA), the multiple sampling on SCA (MS-SCA), and the second attention layer (DPA or transformer).

Model Hyper-parameters

- ***Epoch (80):*** This is the total number of iterations of all the training data in one cycle for training a model.
- ***Batch Size (32):*** Determines the number of images processed in each forward and backward pass during training.
- ***Learning Rate (0.0001):*** Controls the step size at which the model's parameters are updated in response to the estimated error during training.
- ***Optimizer:*** Adam optimizer was used to regularize the model and prevent overfiting.

Results – Evaluation of the model on Test Images (1)

Fig. 3: Predictions and true labels for different classes (person, cat, dog, car, and bicycle).

Results – Evaluation of the model on Test Images (3)

Fig. 4: Predictions and true labels for different classes (cat, dog, car, and bicycle).

Interpretation of Results

Table 1: Performance based on predictions and true labels

		Person	Cat	dog	car	bicycle
Image 1	Predictions	**0.73**	0.16	0.49	0.54	0.65
	True Labels	**1.00**	0.00	0.00	0.00	0.00
Image 2	Predictions	**0.73**	0.46	0.67	0.47	0.73
	True Labels	**1.00**	0.00	0.00	0.00	0.00
Image 3	Predictions	0.72	**0.23**	0.51	0.45	0.62
	True Labels	0.00	**1.00**	0.00	0.00	0.00
Image 4	Predictions	**0.54**	0.29	0.89	**0.48**	**0.89**
	True Labels	**1.00**	0.00	0.00	**1.00**	**1.00**
Image 5	Predictions	0.74	0.25	0.53	0.48	**0.95**
	True Labels	1.00	0.00	0.00	0.00	**0.00**

The IDA model achieving a mAP of **48.6%**.

Summary and Conclusion

This presentation explored how causality can enhance image processing by helping models distinguish true object relationships from misleading contextual elements.

These are key points to note:

- IDA addresses contextual bias using causal inference to improve visual recognition tasks.
- By applying causality, IDA reduces predictions influenced by irrelevant contextual elements, resulting in accurate output.
- Enhances model robustness by focusing attention on the right object features.
- Extend IDA to video recognition and other high-dimensional tasks.

Thanks for listening

References

1. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems (NeurIPS)*, 5998-6008.
3. Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press.
4. Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *European Conference on Computer Vision (ECCV)*, 740-755.