ProbMCL: Simple Probabilistic Contrastive Learning for Multi-label Visual Classification

Ahmad Sajedi [1, 2] Samir Khaki [1] Yuri A. Lawryshyn [1, 2] Konstantinos N. Plataniotis [1]

[1]University of Toronto, Canada [2]Centre for Management of Technology and Entrepreneurship (CMTE), Canada

Introduction and Motivation

Multi-label Visual Classification

A fundamental computer vision task that assigns multiple labels to an input image.

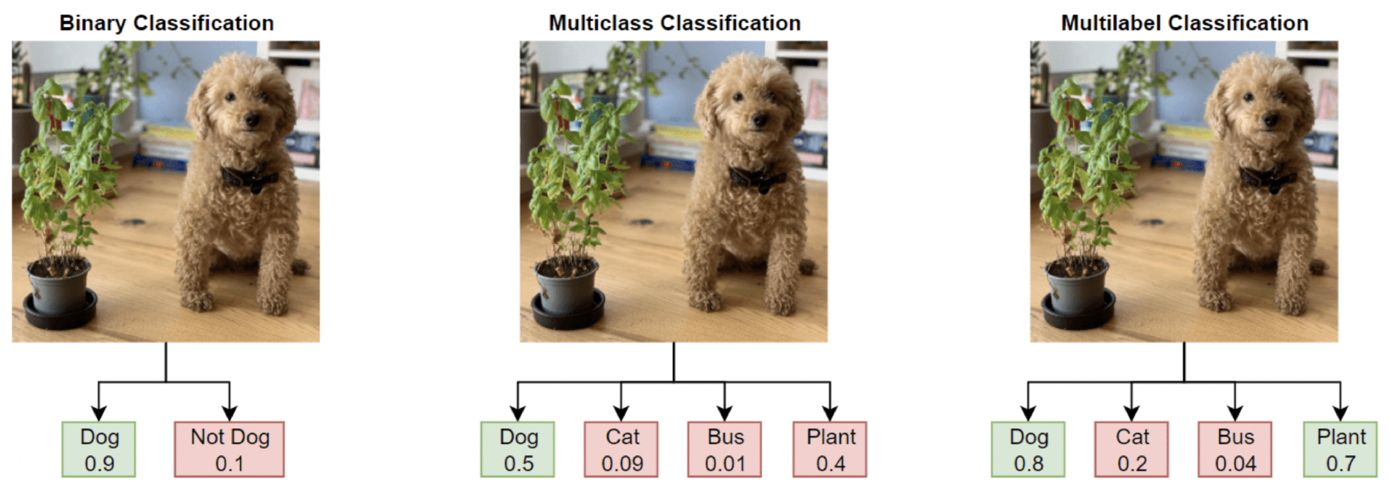


Figure 1. Three visual classification tasks. Multi-label classification assigns multiple labels to an image.

Contrastive Learning

A learning algorithm to extract meaningful representations by contrasting positive and negative pairs of images.

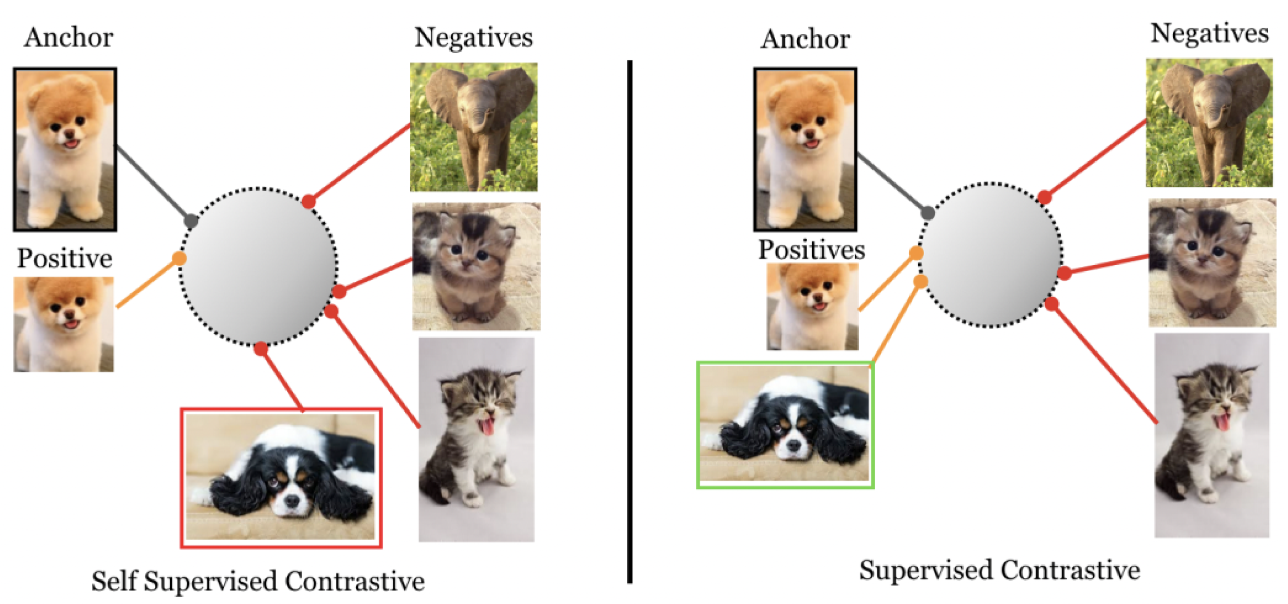


Figure 2. Figure by Khosla et al., Supervised Contrastive Learning, NeurIPS, 2020.

Motivation

- Multi-label classification methods often fail to fully capture label correlations or impose constraints on the learning process through resource-intensive components.
- Existing contrastive learning methods are designed for single-label tasks and lack the necessary smoothness and structures to discern the encoder network's epistemic uncertainty.

Research Question

This study aims to answer the following research question: *Can we develop a simple yet effective contrastive learning algorithm that captures label dependencies and epistemic uncertainty in* ***multi-label classification tasks*** *at a low training cost?*

Contributions

Keeping the research question in mind, we introduce a novel contrastive learning framework designed to overcome existing limitations and enable fast multi-label representation learning for visual classification tasks. Our contributions can be summarized as follows:

- We propose **supervised probabilistic contrastive learning** to efficiently capture label dependencies in multi-label image classification tasks. Our loss function allows for the removal of heavy-duty label correlation modules while achieving optimal performance.
- We integrate a mixture density network into contrastive learning to generate **mixtures of Gaussian** and improve representation learning by estimating feature encoder epistemic uncertainty.
- We employ our pipeline in the computer vision and computational pathology domains to showcase its effectiveness for multi-label image classification across different applications.

Probabilistic Multi-label Contrastive Learning (ProbMCL)

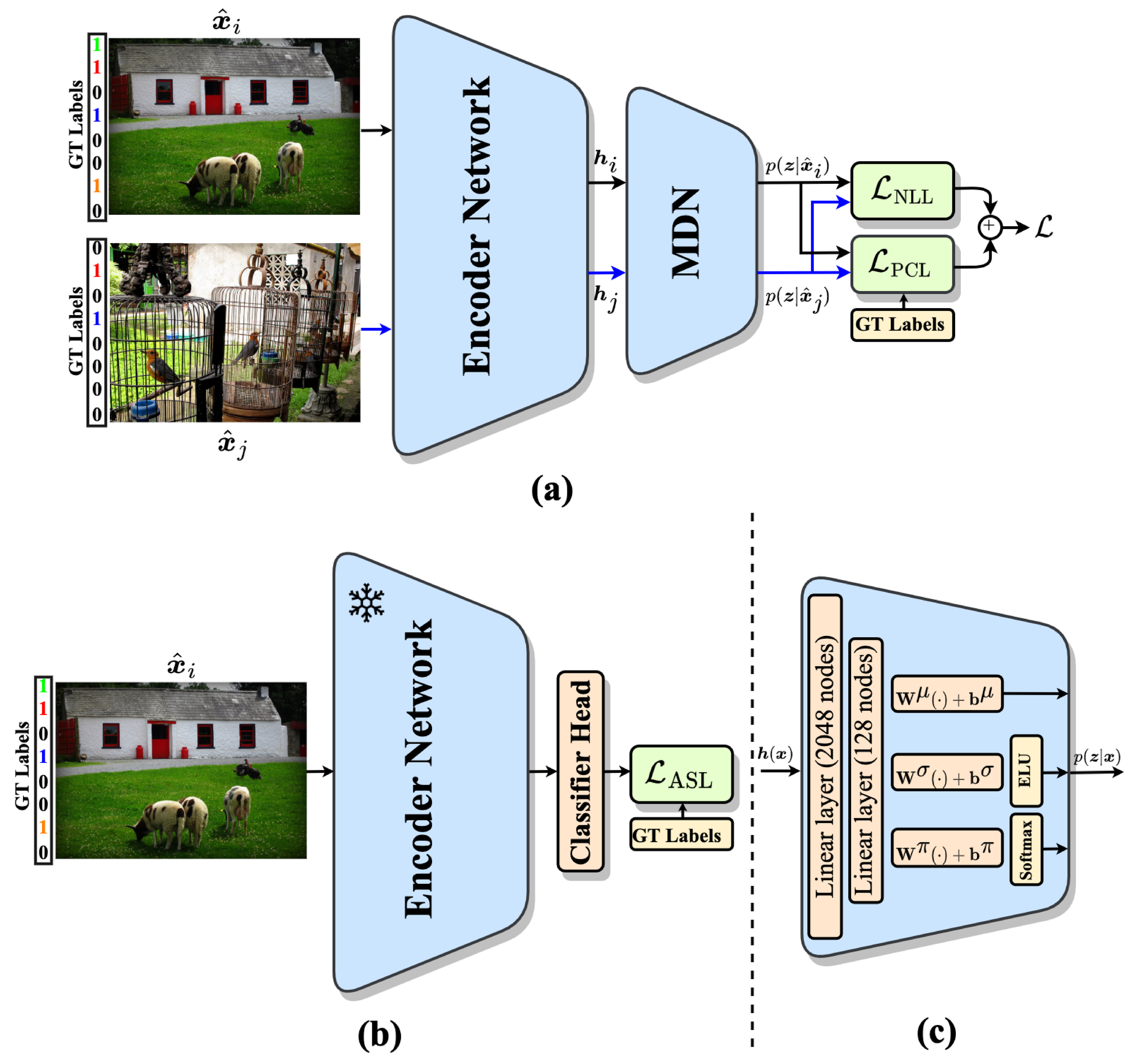


Figure 3. Illustration of the ProbMCL framework in (a) the contrastive stage and (b) the classification stage. In the classification stage, the MDN is discarded and the trained encoder is retained. (c) The internal architecture of the Mixture Density Network (MDN).

Overall Performances on Benchmark Datasets

Table 1. Comparisons with prior methods on MS-COCO. The upper and lower blocks correspond to ResNet-50 and ResNet-101 based models with image resolutions of 224 and 448, respectively. **Bold entries** are the best results.

Method	mAP	CP	CR	CF1	OP	OR	OF1
SRN	77.1	81.6	65.4	71.2	82.7	69.9	75.8
KMCL	82.1	84.1	72.0	77.6	85.0	76.1	80.1
ProbMCL	**82.8**	**85.0**	**72.5**	**78.2**	**85.3**	**76.9**	**80.9**
ML-GCN	83.0	85.1	72.0	78.0	85.8	75.4	80.3
TDRG	84.6	86.0	73.1	79.0	86.6	76.4	81.2
CSRA	83.5	84.1	72.5	77.9	85.6	75.7	80.3
KMCL	88.6	87.7	81.6	84.5	86.3	83.6	84.9
ASL	88.4	85.0	**81.9**	83.4	85.2	84.2	84.7
ProbMCL	**89.1**	**88.5**	**81.9**	**85.1**	**86.7**	**84.3**	**85.5**

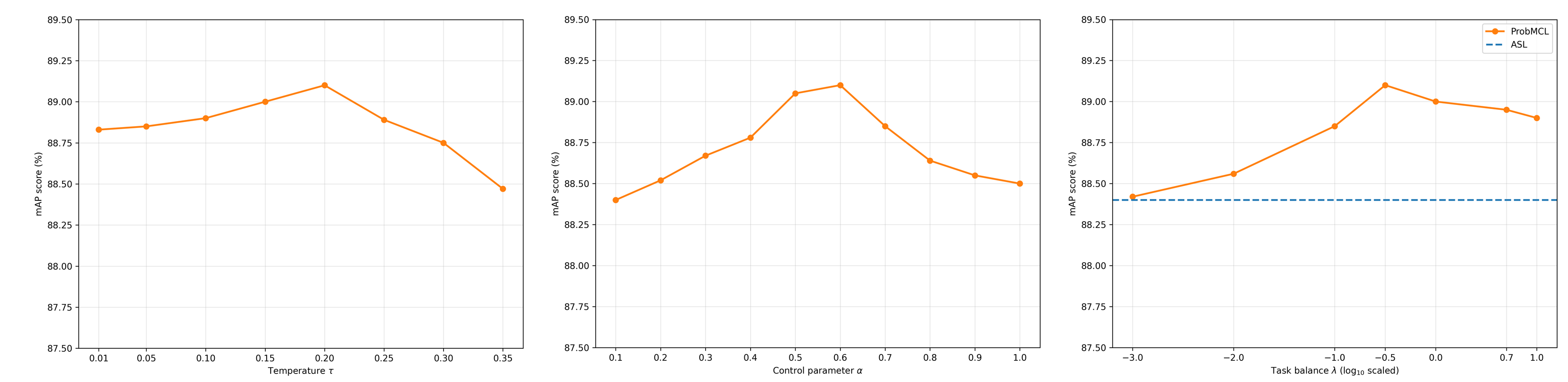
Table 2. Comparisons with state-of-the-art methods on the ADP dataset. **Bold entries** are the best results.

Method	mAP	CP	CR	CF1	OP	OR	OF1
ML-GCN	94.9	91.8	87.0	89.3	92.0	86.9	89.7
TDRG	95.5	**94.6**	84.8	89.4	**94.3**	86.2	90.5
CSRA	96.1	93.1	88.6	90.8	93.0	89.7	91.7
KMCL	96.5	92.6	92.0	92.3	92.7	92.9	92.8
ASL	96.1	93.1	88.6	90.8	92.1	90.7	91.4
ProbMCL	**96.9**	93.0	**92.7**	**92.8**	92.9	**93.3**	**93.1**

Table 3. Computational training cost comparison with prior methods on the MS-COCO dataset. **Bold entries** are the best results.

Method	*ML-GCN*	*TDRG*	*CSRA*	*ProbMCL*
Parameters (M)	44.90	75.20	42.52	**42.23**
GMAC	31.39	64.40	31.39	**29.65**

Ablation Study on Loss Hyperparameters



((a)) Ablation of temperature τ. ((b)) Ablation of control parameter α. ((c)) Ablation of task balance λ.

Figure 4. The effect of loss hyperparameters on the mAP score (%) for the MS-COCO dataset.

Grad-CAM Visualizations

- Superior ability to differentiate dissimilar objects (person and horse)
- Better detection of small objects by capturing label correlation and uncertainty within the novel representations (Person and Blood (H) classes)

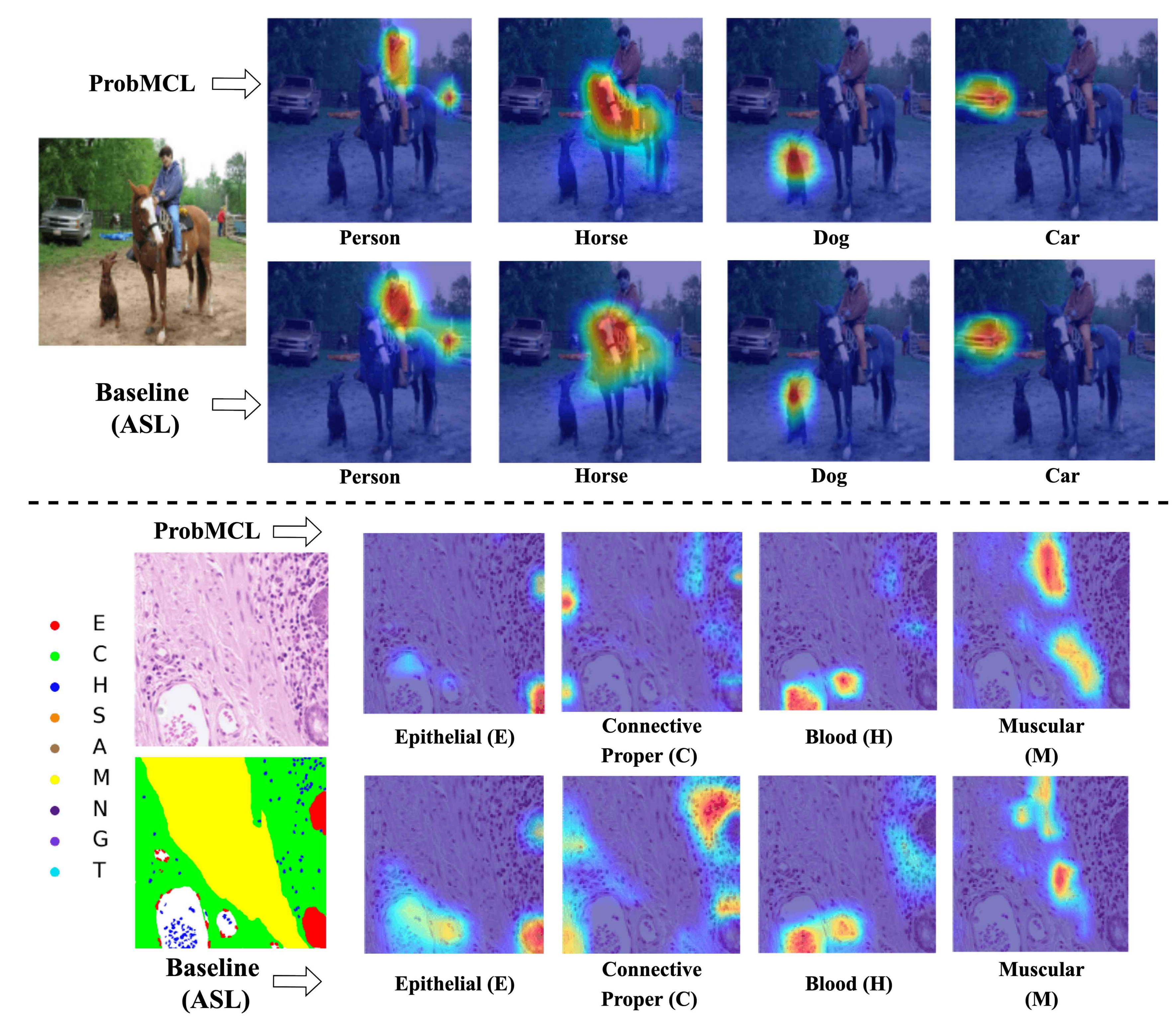


Figure 5. Visualization analyses of baseline (ASL) and the proposed method across MS-COCO (top) and Atlas of Digital Pathology (bottom) datasets.