Dealing with confusing samples into learning based model applied to image classification

Jiarui Xie, Thierry Chateau, Violaine Antoine

Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes

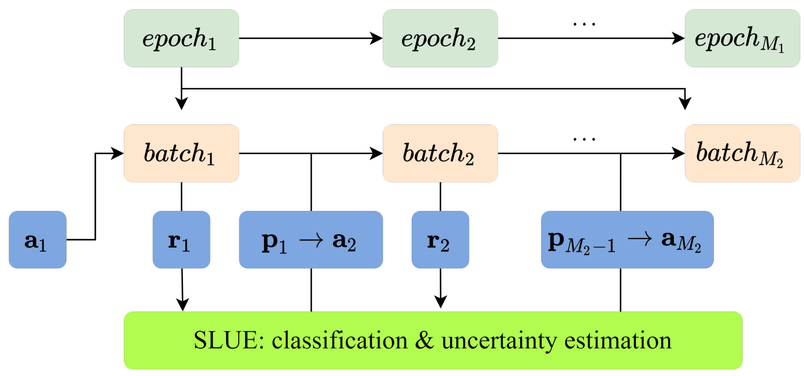
Introduction

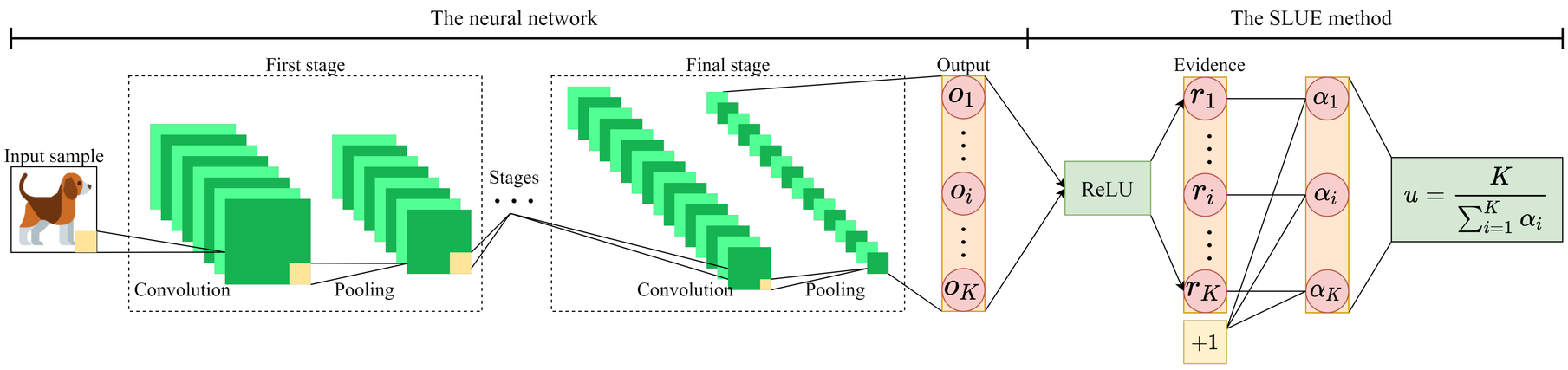
1. **Motivation** The precise classification algorithms suffer from misclassification when encountering confusing samples. Consequently, the ability to detect confusing samples is a guarantee of models' robustness.
2. **Types of the confusing sample**
   - ▷ The out-of-domain sample locates at a large distance from the training dataset or is not represented in the training dataset.
   - ▷ The imprecise sample locates at the intersection of several classes.
3. **Dealing with confusing samples**
   - ▷ Uncertainty estimation represents methods that can calculate an uncertainty value for an input sample. The input sample will be rejected if the obtained uncertainty is bigger than a predefined threshold.
   - ▷ Partial classification represents methods that can calculate the belief or utility for class subsets. Then a class subset is selected as the prediction based on different decision-making strategies.

SLUE: A subjective-logic-based uncertainty estimation mechanism [1]

1. **Motivation** The disadvantage of the existing method (evidential deep learning) [3] is the overlook of base rates.
2. **Contribution** The SLUE method aims at the usage of base rates to guide the training process in the desired direction and fulfill uncertainty estimation.
3. **The framework of the SLUE method.**

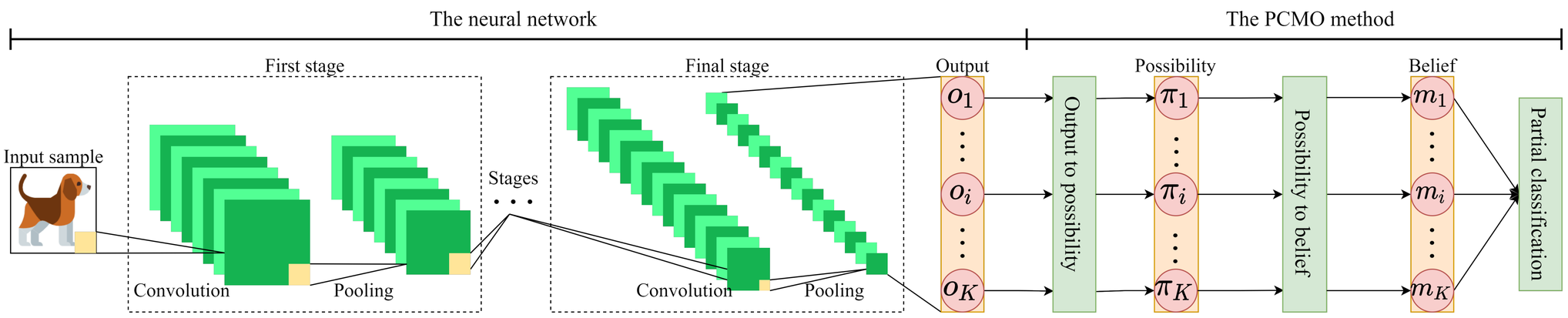
Symbol	Meaning
M_1	the number of epochs
M_2	the number of batches
$\mathbf{a}$	the base rate
$\mathbf{r}$	the evidence
$\mathbf{p}$	the probability
$\boldsymbol{\alpha}$	the Dirichlet distribution parameter





PCMO: Partial Classification from CNN-Based Model Outputs [2]

1. **Motivation** A classification system should provide a potential prediction subset instead of always making rejections. The partial classification method should be efficient and encapsulated as an auxiliary module that can be integrated on the top of existing models.
2. **Contribution**
   - ▷ The PCMO method is fulfilled only based on model outputs that can be applied to a pre-trained neural network without demands to retrain the model or conduct any further modifications.
   - ▷ A novel and reasonable transformation from model outputs to possibility distribution is proposed by considering the $\log$ function.
3. **The framework of the PCMO method.**



References

[1] Jiarui Xie, Violaine Antoine, Thierry Chateau. Partial classification based on pre-trained CNN-based model outputs. International Conference on Neural Information Processing (ICONIP) 2021.
[2] Jiarui Xie, Thierry Chateau, Violaine Antoine. A subjective-logic-based model uncertainty estimation mechanism for out-of-domain detection. International Joint Conference on Neural Networks (IJCNN) 2021.
[3] Sensoy, Murat, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. Advances in Neural Information Processing Systems (NIPS) 2018.