Flexible Visual Recognition by Evidential Modeling of Confusion and Ignorance

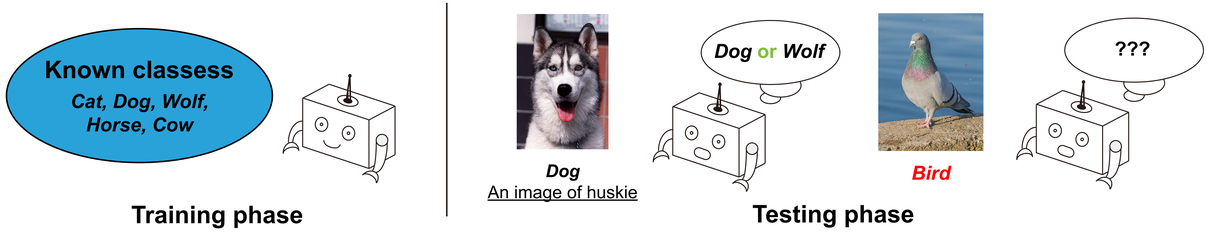
Lei Fan[1], Bo Liu[2], Haoxiang Li[2], Ying Wu[1] and Gang Hua[2]
[1] Northwestern University [2] Wormpex AI Research
leifan@u.northwestern.edu, yingwu@northwestern.edu, {richardboliu, lhxustcer, ganghua}@gmail.com

ICCV23 PARIS

Motivation

In real-world scenarios, visual recognition systems could **fail under two major causes**.

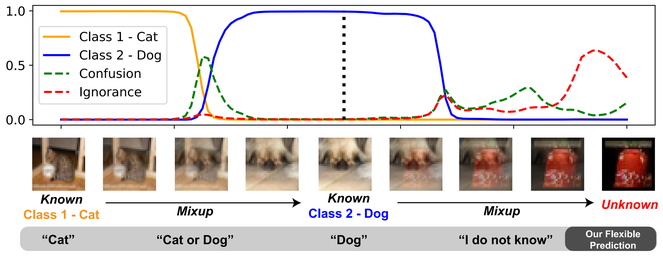
1. **Misclassification between known classes** (left part of the testing phase).
2. **Misbehavior on unknown-class images** (right part of the testing phase).



Two different kinds of uncertainties happen in the above two failure cases. While most previous methods could quantify one whole uncertainty term, we want to model both **Confusion** (left) and **Ignorance** (right) for each sample, separately.

Flexible Visual Recognition

To better exhibit and evaluate the capability of modeling confusion and ignorance, we propose a novel task named '**Flexible Visual Recognition**'.



Classification of the proposed approach on images interpolated from a known-known-unknown triplet.

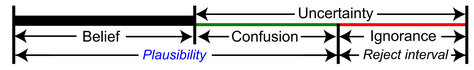
Ignorance reflects the lack of evidence, whereas **confusion** is caused by conflicting evidence, evidence that fails to provide discrimination between specific classes.

A flexible visual recognition system could provide combined predictions when having large confusion and reject making predictions for unknown-class samples.

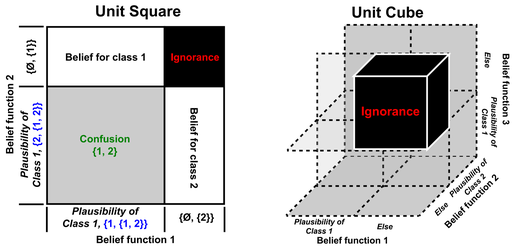
1. Dynamically predict multiple classes when they are unconfident between choices.
2. Reject making predictions when the input is entirely out of the training distribution.

Specifically, we aim to provide a classification model $\mathcal{M}(\cdot)$ that could deliver adaptive predictive sets. For a K-class classification problem, we fomalize the predictive set on image $\mathbf{x}$ as $\{y_1, \ldots, y_k\} = \mathcal{M}(\mathbf{x})$. k obeys $0 \leq k \leq K$. Therefore, k = 0 means the recognition system rejects making a prediction, and the true label $\mathbf{y}$ is supposed to be contained in the predictive set when $\mathbf{x}$ is from known classes.

Method



Relation between different masses constituting the final set of opinions towards a hypothesis.



Two graphical demonstrations of evidence combination on 2- and 3-class classification task.

We develop our method under **the theory of Subjective Logic. Confusion** is the shared evidence contributing to multiple categories while not discriminative between them, while **ignorance** is completely missing evidence.

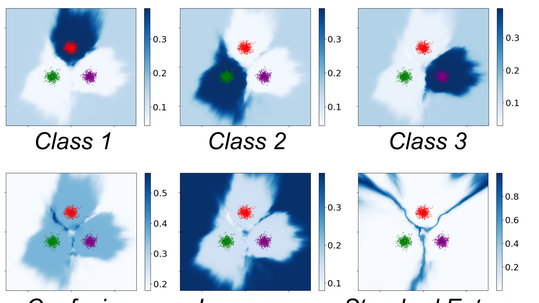
Uncertainty = Confusion + Ignorance

Uncertainty + Beliefs belong to single class = 1

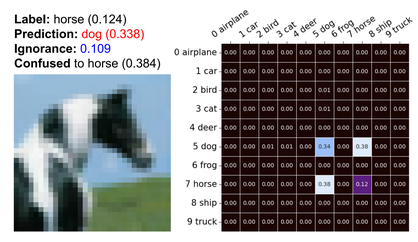
Confusion = Beliefs on non-singleton sets of classes

During training, the model learns the Dirichlet prior placed on singleton classes. Confusion and ignorance could then be obtained through the evidence combination theory.

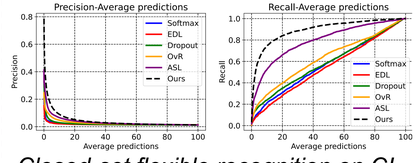
Results



A 3-class classification problem. The Gaussian-distributed training data are depicted with dots, while the background color indicates the predicted value of the corresponding location. We provide our model's estimated **single-class belief, confusion** and **ignorance**. The entropy over predictions of a standard net trained with cross-entropy loss is shown in the last subfigure for comparison.



Matrices of confusion of misclassified samples on the CIFAR-10 dataset.



Closed-set flexible recognition on CIFAR-100. Precision and recall of delivering multiple predictions on misclassified samples with respect to the average number of predictions. The intention to predict extra classes is dependent on the confusion term between classes.

The proposed method could achieve per-sample confusion and ignorance quantifications. The diagonal of each matrix is set to the singleton belief of each class.The total ignorance is demonstratd in the caption.

Our method could achieve sample-wise comparable confusion so that the model knows when to make multiple predictions.

Please refer to our paper for quantitative comparisons.