UNIVERSITÄT BERN

ARTORG CENTER
BIOMEDICAL ENGINEERING RESEARCH

goFOOD™
ARTIFICIAL INTELLIGENCE MEETS NUTRITION

MADiMa 2020 VIRTUAL

Food Recognition in the Presence of Label Noise

Ioannis Papathanail[1], Ya Lu[1], Arindam Ghosh[2], Stavroula Mougiakakou[1]

[1] ARTORG Center for Biomedical Engineering Research, University of Bern

[2] Oviva S.A., Zurich, Switzerland

Contents

- Introduction
- Method
- Database
- Experimental Results
- Conclusion

Introduction - Motivation

- Adhering to a healthy diet can not only prevent overweight and obesity but also lower the risk of numerous chronic diseases.
- Recently, computer vision dietary based assessment has replaced traditional methods [1] like food diaries and 24-hour recall to monitor a person's dietary habits.
- Computer vision methods depend on large-scale databases with clean annotations. However, images taken from real end-users tend to contain noisy labels.
- Constructing a clean dataset can prove very costly and time consuming, therefore, a noisy dataset is often preferred instead.

[1] Thompson, F. E., Subar, A. F.: Dietary assessment methodology. In *Nutrition in the Prevention and Treatment of Disease*, pp. 5-48, Academic Press (2017).

Introduction - Goal

Propose an effective approach to deal with noisy labels in a multi-label food recognition problem, where:

- The addition of a Noise Layer (NL) [1] can improve the results on any baseline image classification model (BM) (e.g., ResNet, InceptionV3).
- The NL does not increase the computation time, since it is used only for the training procedure and it is removed afterwards.
- The method does not rely on a subset of images with clean labels that can assist in the training process.

[1] Goldberger, J., Ben-Reuven E.: Training deep neural-networks using a noise adaptation layer, (2016).

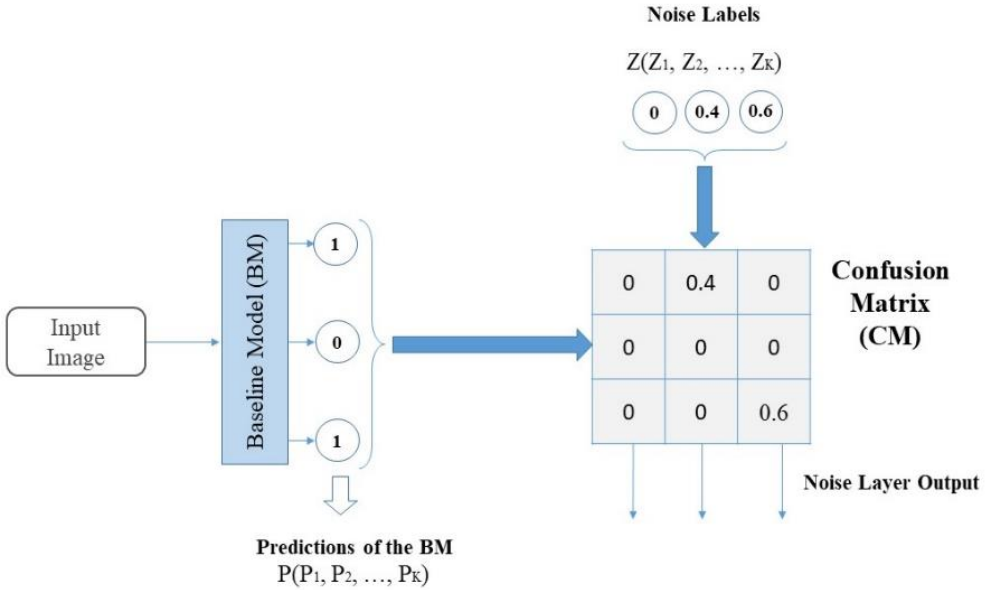
Method (1/3)

- First, train a baseline image classification model (BM) on the noisy dataset
- Build a confusion matrix (CM) based on the predictions of the BM on the training set and the noisy labels
- Build a Noise Layer on top of the BM. This is the full model (FM)
- Retrain the FM on the training set.
- Make the predictions on the clean testing set, after removing the NL that was used to learn the noise distribution in the training set.

Method (2/3)

Confusion Matrix (CM) Building:

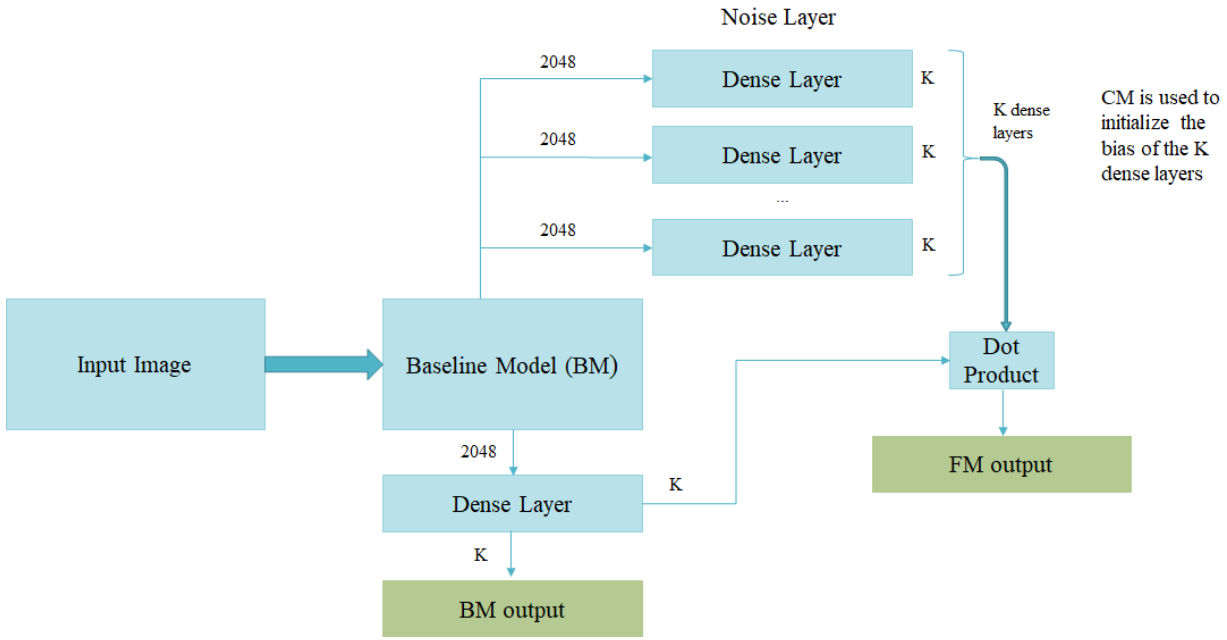
- Rows: Predictions *P* of the BM for an image, after applying a threshold
- Columns: The average of the annotations for each class for the image
- The elements of the CM are updated for each new image



Method (3/3)

Full Model (FM):

- For each of the K outputs of the BM, add a Dense Layer with K units.
- Initialize the k^{th} Dense Layer with bias equal to the k^{th} row of the CM.
- The output of the FM is the dot product of the BM output and the outputs of the Dense Layers that constitute the Noise Layer.

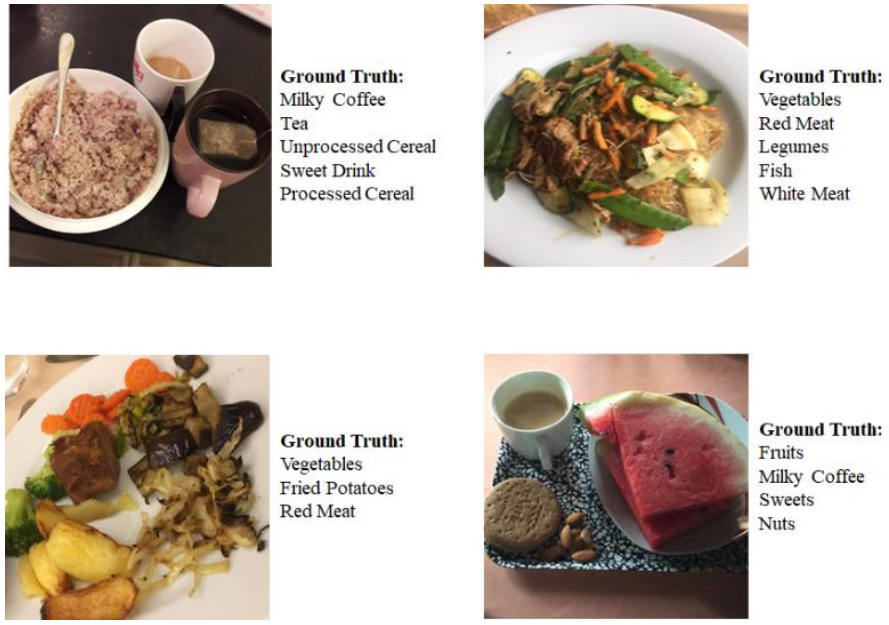


Database

Medipiatto Dataset:

- Images taken under free living conditions
- 31 food categories
- Annotated by 5 inexperienced annotators, testing set labels were corrected by an experienced dietitian
- 5485 images for training and 293 images for testing
- Average of 3.74 food items in each image

Example images of the training set (upper row) and the testing set (lower row) of the database along with their annotations



Experimental Results (1/3)

Evaluation Metrics:

- Mean Average Precision (mAP):

$$mAP = \frac{1}{K}\sum_{k=1}^{K} mean\left(\max\left(P_{\mathrm{R}}^{k}\right)\right)$$

- Per-class Average Precision (AP):

$$AP_k = mean\left(\max\left(P_{\mathrm{R}}^{k}\right)\right)$$

where $\max\left(P_{\mathrm{R}}^{k}\right)$ is the maximum precision at each recall value, for category k.

Experimental Results (2/3)

Comparison of mAP between the FM and the BM for the InceptionV3 and the ResNet-101 architecture.

Model Architecture	**mAP**
InceptionV3	0.416
InceptionV3 with NL	0.499
ResNet-101	0.466
ResNet-101 with NL	**0.507**

Experimental Results (3/3)

Samples of images in the testing set along with predictions from the BM and FM using the ResNet-101 architecture. The categories appear in green, red and red with grey background for correct, wrong and missing predictions, respectively.

Ground Truth:
Vegetables
Cheese
Red Meat
BM predictions:
Vegetables
Red Meat
Cheese
FM predictions:
Vegetables
Cheese
Red Meat

Ground Truth:
Vegetables
Non-fried Potatoes
BM predictions:
Vegetables
Fried Potatoes
Breaded Food
Non-fried Potatoes
FM predictions:
Vegetables
Non-fried Potatoes
Red Meat

Ground Truth:
Red Meat
Eggs
White Bread
BM predictions:
Eggs
Red Meat
White Bread
FM predictions:
Eggs
White Bread
Red Meat

Ground Truth:
Yoghurt
Cheese
White Bread
Sweet Drink
Coffee
Processed Cereal
BM predictions:
Yoghurt
Cheese
Tea
White Bread
Sweet Drink
Coffee
Processed Cereal
FM predictions:
Yoghurt
Cheese
Coffee
White Bread
Sweet Drink
Processed Cereal

Conclusions

- In this work we proposed a method to combat label noise in a multi-label food recognition problem.
- The FM that contains the BM and the NL outperforms the BM for different network architectures while not increasing the computation time.
- Future work includes the evaluation of the proposed method on much larger datasets.

Thank you for the attention!
Questions?

ioannis.papathanail@artorg.unibe.ch
ya.lu@artorg.unibe.ch
arindam.ghosh@oviva.com
stavroula.mougiakakou@artorg.unibe.ch