Uncertainty-aware Data Augmentation for Food Recognition

Eduardo Aguilar, Bhalaji Nagarajan, Rupali Khatun, Marc Bolaños, Petia Radeva.

Presented by: Eduardo Aguilar

Can there be data that require more attention than others?

- Within a dataset, there could be more and less complex classes.
- Some images could be more beneficial for the classification than others.

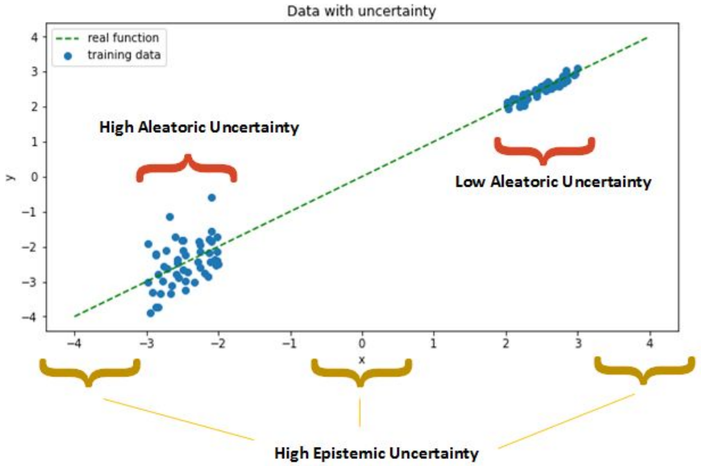
More emphasis should be placed on these data.

Hypothesis: Estimating the uncertainty can help us decide the most appropriate data to perform **additional** data augmentation methods.

Uncertainty

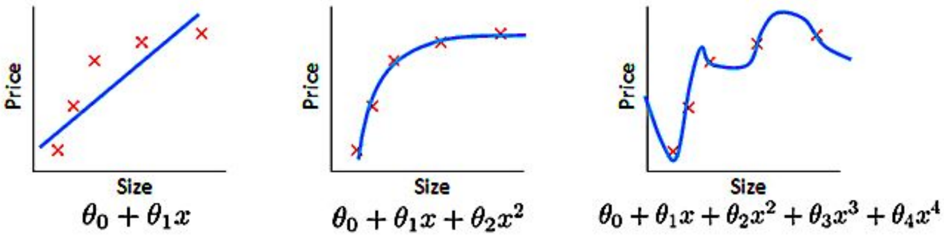
Aleatoric – captures the noise inherent in the observations

- Heteroscedastic – data-dependent
- Homoscedastic – constant for different data points, but can be task-dependent.



Epistemic – model uncertainty

- Can be explained away given enough data
- Uncertainty about the model parameters
- Uncertainty about the model structure



Samples of the Smallest and Largest EU within the same class of Dish

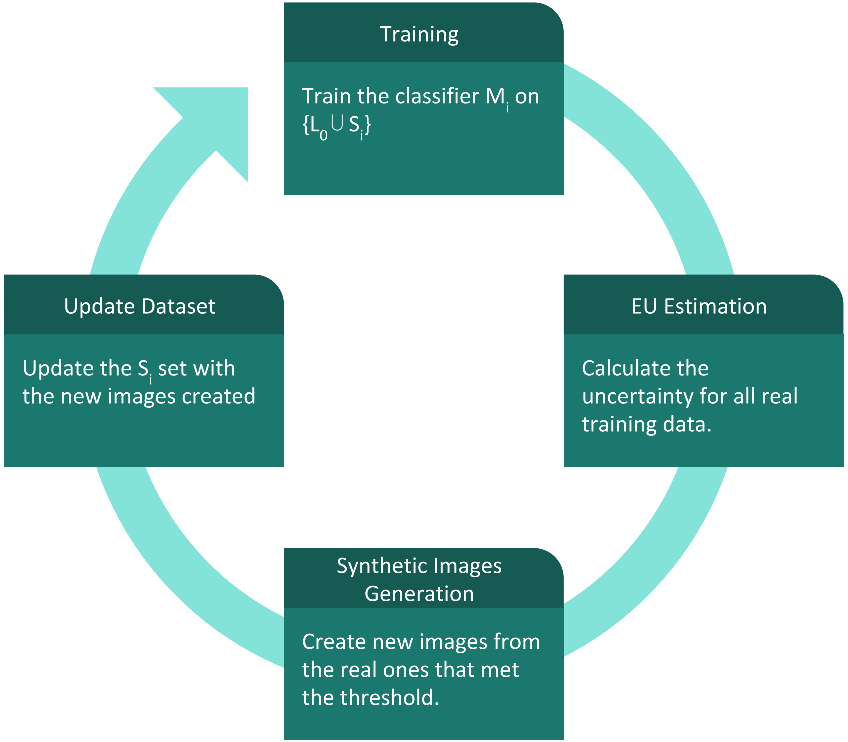
Caesar Salad

Ravioli

Steak

Tacos

UDA Procedure



Algorithm 1: UDA procedure

input: Labeled Data L_0, Synthetic Data $S_i \leftarrow \emptyset$, Lower Threshold T_l, Upper Threshold T_u, Generator G;

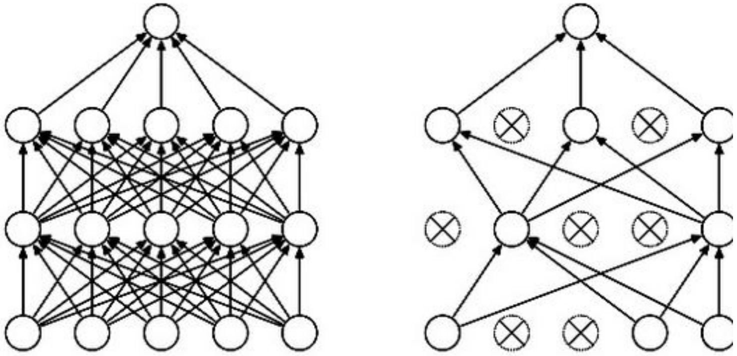
while *we are not satisfied with the performance* **do**
    Train the classifier M_i on $\{L_0 \cup S_i\}$;
    for $x_j \in L_0$ **do**
        Calculate the $EU(x_j)$;
        if $T_l < EU(x_j) < T_u$ **then**
            Create the synthetic image x_j^* with G;
            Update $S_i \leftarrow S_i \cup \{x_j^*\}$;
        end
    end
end

EU Estimation: MC-Dropout

1. Infer y|x multiple times, each time sample a different set of nodes to drop out.
2. Average the predictions to get the final prediction E(y|x).
3. Calculate the sample variance of the predictions.

$$EU(x_t) = -\sum_{c=1}^{C} \overline{p(y_c = \hat{y}_c|x_t)} \ln(\overline{p(y_c = \hat{y}_c|x_t)}),$$

$$\overline{p(y_c = \hat{y}_c|x)} = \frac{1}{K}\sum_{k=1}^{K} p(y_c^k = \hat{y_c^k}|x).$$



(a) Standard Neural Net (b) After applying dropout.

Synthetic Images

The synthetic image generation is performed by applying various transformations on the real images, such as **rotation, width and height shift, shear, zoom**, etc.

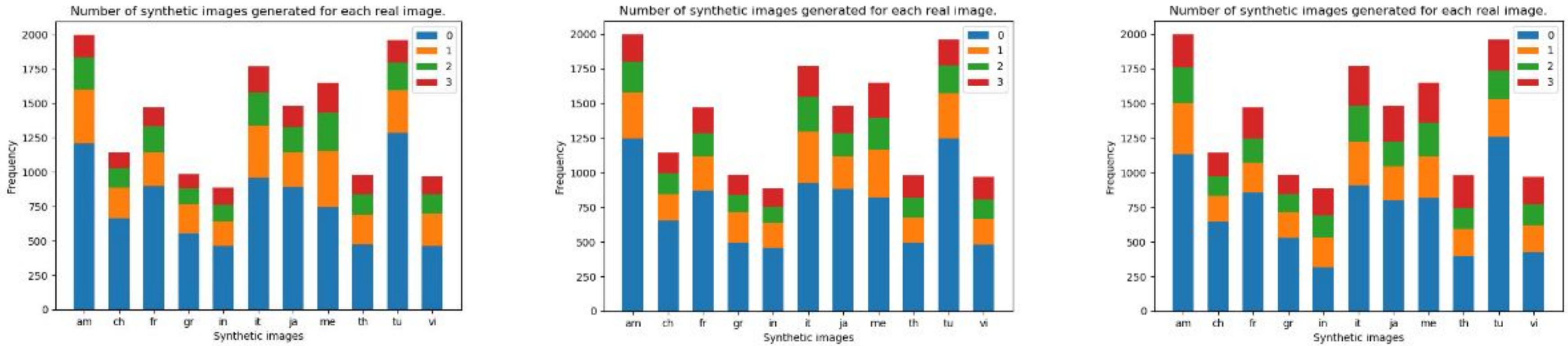
Validation: MAFood121
A new Multi-Attributes Food Datasets

A public multi-attribute food datasets with:

- 121 dishes within the 11 most popular cuisines.
- More than 20.000 images.
- Annotations of 20.000 images for 3 food-related tasks.

E. Aguilar, M. Bolaños, and P. Radeva. Regularized Uncertainty-based Multi-task Learning Model for Food Analysis. JVCI (2019)

Validation



(a) ResNet50_DO (b) InceptionV3_DO (c) DenseNet169_DO

Fig. 4. Number of synthetic images generated after the third training cycle. The blue color represents the real images that were not used to generate the synthetics. As for the colors orange, green and red, these illustrate the number of real images used, one, two or three times to generate synthetic images.

Results

TABLE II
RESULTS OBTAINED ON THE TEST SETS IN TERMS OF R_{micro}.

Method	American	Chinese	French	Greek	Indian	Italian	Japanese	Mexican	Thai	Turkish	Vietnamese
DenseNet169_DO (S_1)	87,12%	91,83%	93,56%	91,10%	93,68%	86,26%	93,62%	83,07%	84,95%	93,81%	90,67%
DenseNet169_DO (S_2)	88,89%	92,61%	94,24%	93,19%	93,68%	86,49%	93,62%	85,62%	84,95%	93,58%	91,19%
DenseNet169_DO (S_3)	**89,39%**	93,00%	93,22%	92,67%	**93,68%**	87,68%	**94,33%**	86,26%	**86,02%**	**94,03%**	**91,71%**
DenseNet169_DO (S_4)	88,89%	**94,16%**	**94,92%**	**93,72%**	93,16%	**87,91%**	**94,33%**	**86,90%**	**86,02%**	**94,03%**	91,19%

TABLE III
RESULTS OBTAINED ON THE TEST SETS IN TERMS OF R_{macro}.

Method	American	Chinese	French	Greek	Indian	Italian	Japanese	Mexican	Thai	Turkish	Vietnamese
DenseNet169_DO (S_1)	86,93%	91,84%	94,02%	89,60%	93,53%	87,64%	92,75%	82,31%	82,69%	93,68%	90,21%
DenseNet169_DO (S_2)	88,77%	92,08%	93,96%	92,42%	93,51%	87,19%	93,19%	85,55%	83,84%	93,55%	91,31%
DenseNet169_DO (S_3)	**89,26%**	92,69%	93,24%	91,59%	**93,77%**	88,83%	93,81%	86,73%	83,99%	93,96%	**91,50%**
DenseNet169_DO (S_4)	88,98%	**92,89%**	**95,78%**	**92,99%**	92,79%	**89,42%**	**94,04%**	**86,97%**	**84,08%**	**94,00%**	91,18%

S1: Training with real images without applying any typeof data augmentation.
S2: Training with standard online data augmentation, such as random crops and horizontal flips, applied on real images only.
S3: Training with standard online data augmentation,applied on a dataset consisting of both synthetic images(one for each real image) and real images.
S4: Training with standard online data augmentation,applied on a dataset generated by our UDA method.

Conclusions

- A novel method for uncertainty-aware data augmentation that follows an active learning framework.
- We validated our approach on eleven subsets of data from the public food dataset MAFood-121.
- It is not necessary to generate data-augmented images for all the samples.
- Our method contributes to getting more balanced classification in the unbalanced dataset.