Uncertainty-Aware Data Augmentation for FoodRecognition

Eduardo Aguilar, Bhalaji Nagarajan, Rupali Khantun, Marc Bolaños and Petia Radeva

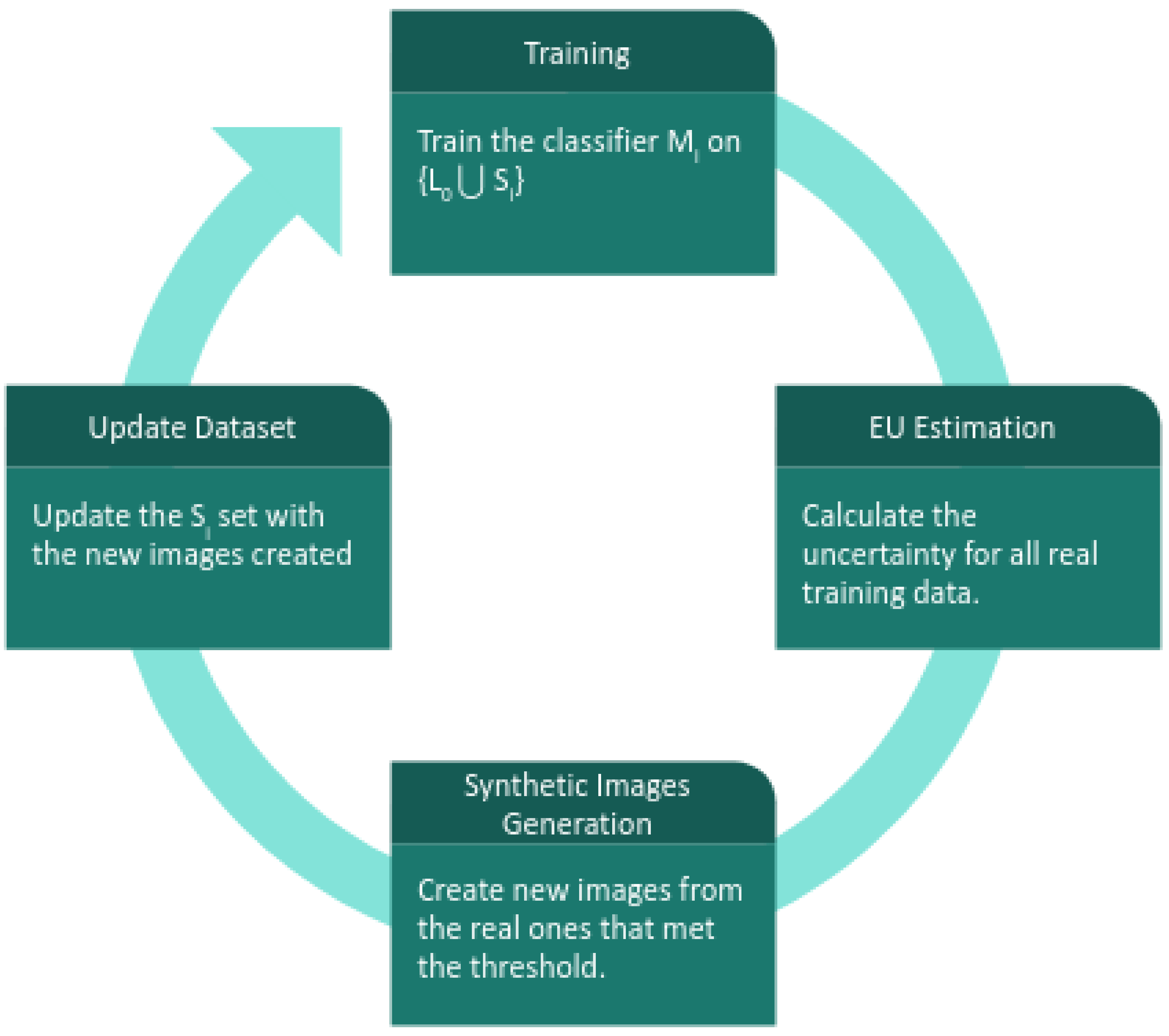
25th International Conference on Pattern Recognition (ICPR 2020)

Motivation

1) Common data augmentation techniques apply tranformations in whole dataset. However, do we really need to increase all the data interchangeably?
2) When dealing with fine-grained object recognition, such as food recognition, we often find that some images tend to be more complex to classify than others. So how can we discover which images are probably more complex?
3) Emphasizing data augmentation in hard samples could help the model understand its distinctive characteristics.

Goal: To provide a data augmentation procedure that focus on complex images to guide the model learning.

Method



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

For the validation, the Epistemic Uncertainty was estimated based on the MC-Dropout method, which can be expressed as follows:

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

where

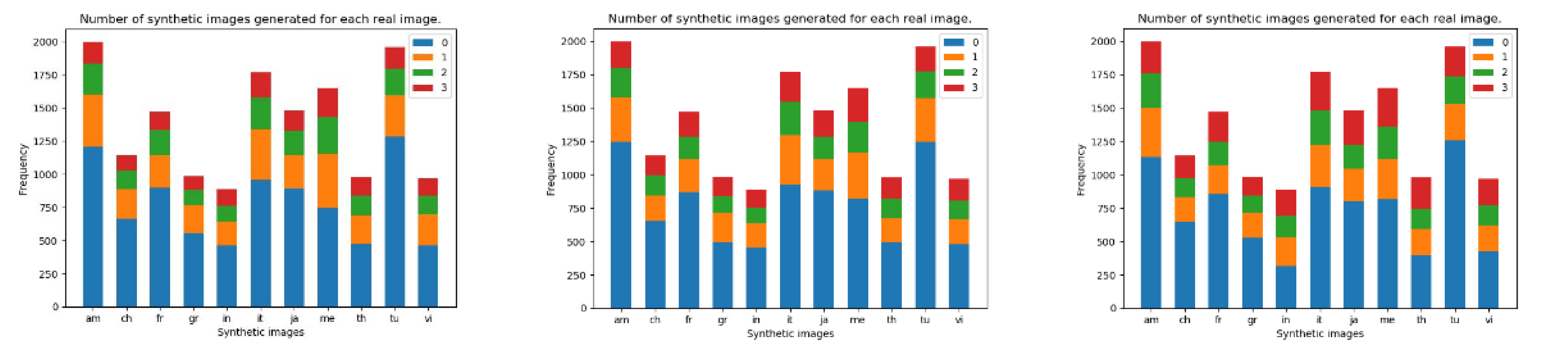
$$\overline{p(y_c = \hat{y}_c|x)} = \frac{1}{K}\sum_{k=1}^{K} p(y_c^k = \hat{y_c^k}|x).$$

*K denotes the number of MC-dropout iterations.

MAFood-121

SL tasks:
Dish and cuisine
ML task:
Food categories
#Images:
21.175
splits (train, val, test):
72,5%, 12,5%, 15%.

Results



(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.

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
ResNet50_DO (S_1)	82,07%	89,49%	89,49%	90,05%	87,37%	79,15%	89,36%	79,23%	76,88%	91,37%	85,49%
ResNet50_DO (S_2)	83,84%	89,49%	91,53%	90,05%	92,63%	81,75%	81,84%	79,87%	**82,26%**	91,59%	87,59%
ResNet50_DO (S_3)	84,09%	91,44%	**92,88%**	91,10%	**93,16%**	82,70%	**92,20%**	**82,11%**	**82,26%**	91,81%	88,60%
ResNet50_DO (S_4)	**84,34%**	**91,83%**	92,20%	**92,15%**	92,11%	**83,65%**	91,49%	81,15%	81,18%	**92,04%**	**90,16%**
InceptionV3_DO (S_1)	84,60%	90,27%	91,86%	91,10%	92,11%	83,65%	**92,55%**	81,47%	81,72%	91,15%	86,051%
InceptionV3_DO (S_2)	87,12%	91,05%	92,54%	92,15%	92,11%	**86,73%**	**92,55%**	82,43%	82,80%	92,92%	86,53%
InceptionV3_DO (S_3)	**87,88%**	**91,44%**	93,22%	**92,67%**	91,58%	**86,73%**	**92,55%**	83,71%	84,95%	**93,81%**	88,60%
InceptionV3_DO (S_4)	**87,88%**	**91,44%**	**93,56%**	**92,67%**	**94,21%**	86,49%	92,20%	**84,35%**	**85,48%**	93,14%	**89,64%**
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
ResNet50_DO (S_1)	81,99%	87,93%	89,01%	89,12%	87,67%	80,72%	88,08%	79,12%	70,98%	91,44%	84,67%
ResNet50_DO (S_2)	83,69%	90,05%	90,33%	89,34%	92,96%	82,44%	90,85%	80,37%	**79,91%**	91,65%	86,99%
ResNet50_DO (S_3)	84,10%	90,60%	**94,12%**	89,90%	**93,29%**	**84,31%**	**91,20%**	81,64%	79,85%	91,92%	88,14%
ResNet50_DO (S_4)	**84,26%**	**91,17%**	92,54%	**92,11%**	92,41%	84,07%	90,93%	**81,96%**	79,22%	**92,15%**	**89,85%**
InceptionV3_DO (S_1)	84,64%	89,97%	91,21%	89,80%	91,98%	84,43%	91,64%	81,98%	80,75%	91,48%	84,82%
InceptionV3_DO (S_2)	87,16%	**91,27%**	91,88%	90,94%	92,19%	88,15%	92,05%	82,85%	80,85%	92,84%	85,05%
InceptionV3_DO (S_3)	**87,92%**	90,69%	91,78%	91,52%	91,56%	88,23%	**92,50%**	84,14%	**82,98%**	**93,80%**	88,49%
InceptionV3_DO (S_4)	87,83%	91,22%	**92,60%**	**91,93%**	**93,69%**	**88,25%**	91,47%	**84,19%**	82,59%	93,31%	**88,60%**
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%

Acknowledgements

This work was partially funded by TIN2018-095232-B-C21, 2017 SGR 1742, Nestore H2020-EU.3.1.4. No:769643, Validithi, 20141510 (La MaratoTV3) and CERCA Programme/Generalitat de Catalunya. E. Aguilar acknowledges the support of CONICYT Becas Chile. P. Radeva is partially supported by ICREA Academia 2014. We acknowledge the support of NVIDIA Corporation with the donation of GPUs.

Conclusions

- ✓ Novel method for uncertainty-aware data augmentation that follows an active learning framework and take into account the most uncertain images to generate new ones.
- ✓ The method was validated on 11 subsets of MAFood121 and shown that it isn't necessary to generate data-augmented for all samples to improve performance.