DOES THE DATASET MEET YOUR EXPECTATIONS? EXPLAINING SAMPLE REPRESENTATION IN IMAGE DATA

Dhasarathy Parthasarathy[1,2] Anton Johansson[2]

[1]Volvo Group, Sweden
[2]Chalmers University of Technology, Sweden

How can we reason about our datasets?

Ensuring that our training set represents an adequate variety of scenarios is crucial in ensuring the reliability of deep learning systems. With practical datasets typically being high-dimensional and large, explaining the adequacy of sample representation is not straightforward.

- Given a dataset $\mathcal{S} = \{X_i, Y_i\}_{i=1}^{N}$, we approach this problem by considering the annotations associated with each datapoint $Y_i \sim P_{\mathcal{S}}(Y)$
- Further, we *specify* a distribution of annotations $P_{\mathcal{T}}(Y)$, expressing the sample representation that is *expected* in the dataset. Deficiencies in our dataset can then be explained as the mismatch between $P_{\mathcal{T}}(Y)$ and $P_{\mathcal{S}}(Y)$.
- In practice, most datasets are not adequately labeled and $P_{\mathcal{S}}(Y)$ is thus not available. Combining simulation, outlier detection, input attribution, and a new overlap index, we show that it is possible to visualize, quantify and explain sample representation in a comprehensible low-dimensional form, even when annotations are not explicitly available in $\mathcal{S}$.

Explaining sample representation using annotations

To investigate the validity of the proposed method, we will consider a dataset of circles and squares where the true annotations can easily be obtained

$$P_{\mathcal{T}}(Y)$$
Side-length of bounding box: $Y^S \sim \mathcal{U}\{30, 120\}$
Top left corner of bounding box: $Y^2, Y^3 \sim \mathcal{U}\{0, 128 - Y^S\}$
Bottom right corner of bounding box: $Y^4, Y^5 \sim \mathcal{U}\{Y^S, 128\}$
Average pixel brightness: $Y^6 \sim \mathcal{U}\{100, 255\}$

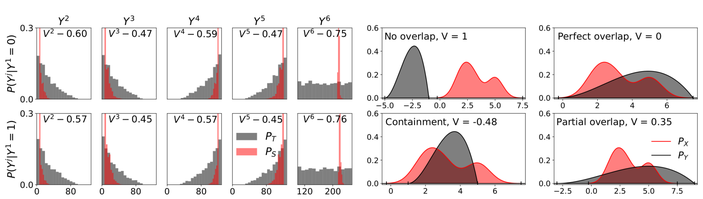
Defining an *overlap index* as

$$R_X = \{x \in \mathbb{R} : P_X(x) > 0\} \tag{1}$$

$$V(P_X, P_Y) = I\,\frac{\lambda(R_X \,\Delta\, R_Y)}{\lambda(R_X \,\cup\, R_Y)}, \quad I = \begin{cases} -1 & R_Y \subset R_X \\ +1 & otherwise \end{cases} \tag{2}$$

$$V^j(P) = V(P_{\mathcal{S}}(Y^j \,|Y^1), P(Y^j \,|Y^1)), \quad j = 2...6 \tag{3}$$

The mismatch between $P_{\mathcal{T}}(Y)$ and $P_{\mathcal{S}}(Y)$ can then be visualized and quantified.

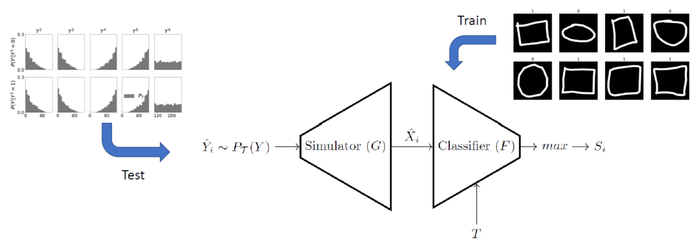


Explaining sample representation using simulation

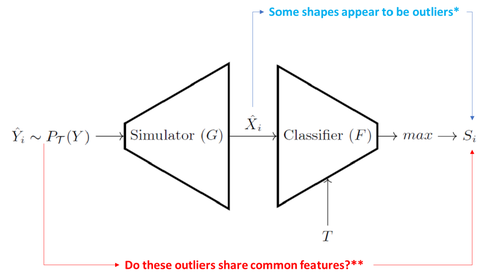
When labelling is not possible, we can recognize that parametric simulation can invert the labelling process. Exploiting this notion, we explain sample selection bias with a 2-step process (i) detecting outlier annotations and (ii) estimating marginal sample representation.

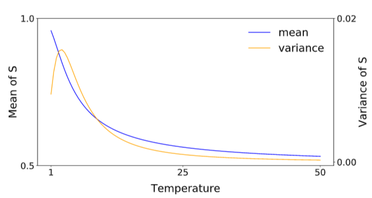
Step 1 - Detecting outlier annotations

We work under the following assumption - *a test annotation $\hat{Y}_i$, that is unlikely to be observed in $\mathcal{S}$, maps to a simulated test sample $\hat{X}_i = G(\hat{Y}_i)$, that appears as an outlier to $\mathcal{S}$.* This converts the problem of exploring sample representation to one of outlier detection.

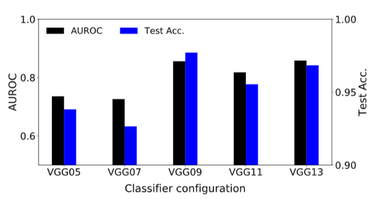


$$S_i = E_{\mathcal{S}}(\hat{Y}_i, F, T) = \max_{k \in K} \frac{exp(F_k(G(\hat{Y}_i))/T)}{\sum_{k \in K} exp(F_k(G(\hat{Y}_i))/T)}, \hat{Y}_i \sim P_{\mathcal{T}}(Y), K = \{0, 1\}$$





(a) Effect of temperature scaling on the distribution of uncertainty scores F=VGG13



(b) AUROC for detecting outlier annotations per classifier at $T = T^*$ and $S^T = 0.7$

Step 2 - Estimating marginal sample representation

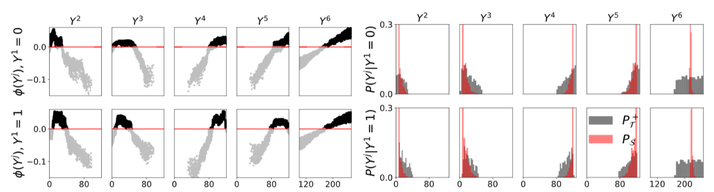
Shapley Additive Explanations (SHAP) enables us to assess individual design concerns like size, position, and brightness.

$$S_i = E_{\mathcal{S}}(\hat{Y}_i, F, T) = \phi^0 + \sum_{j=2}^{6} \phi_i^j \tag{4}$$

This allows us to estimate the marginal sample representation, a proxy for verifying our design concerns.

$$P_{\mathcal{T}}^{+}(Y^j = l \,|Y^1 = k) = \frac{|\{\hat{Y}_i^j : \phi_i^j \geq 0,\ \hat{Y}_i^j \in Y^l\}|}{|\{\hat{Y}_i^j : \phi_i^j \geq 0\}|},\ j = 2...6,\ \hat{Y}_i \in \hat{Y}$$
$$Y^l = \{l - \delta, l + \delta\}, \quad \hat{Y} = \{\hat{Y}_i : \hat{Y}_i^1 = k\},\ k \in K \tag{5}$$

The marginal sample representations $P_{\mathcal{T}}^{+}$ can then act as a substitute for the marginal distributions of $P_{\mathcal{S}}(Y)$.



(a) Sample-representation from SHAP scores **(b)** Marginal sample representation

Conclusion

There is a need to understand the sample representation in the training data to ensure reliability of the trained model. We approach this problem through annotations associated with each data point. On a dataset of circles and squares, we show that through a combination of simulation, outlier detection and input attribution, it is possible to obtain a visualization and quantification of the sample representation without a need for an expensive labelling system.

Acknowledgements

This work was supported by the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation.

References

*Dan Hendrycks and Kevin Gimpel (2016). A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural NetworksCoRR, abs/1610.02136.

**Scott M. Lundberg and Su-In Lee (2017). A Unified Approach to Interpreting Model Predictions. In Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA (pp. 4765–4774).