Explainable AI: Object Recognition With Help From Background

CSS Workshop ICLR2022

Raza Hashmi, Qiang Li

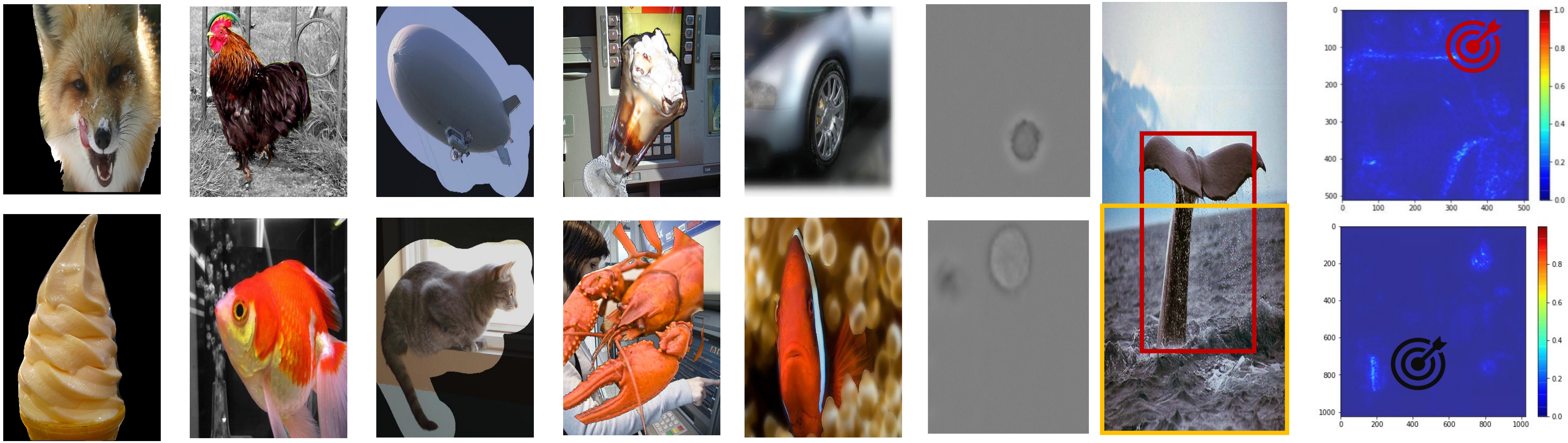


TABLE OF CONTENTS/AGENDA

1) RELATED WORK

i. Xiao et al., ***2021 ICLR NOISE OR SIGNAL: THE ROLE OF IMAGE BACK- GROUNDS IN OBJECT RECOGNITION***

2) INTRODUCTION

3) METHODOLOGY

i. Class independent factors in background
   i. Foreground only.
   ii. Color.
   iii. Blur.
   iv. Size of the background.

ii. Class dependent factors in background
   i. Highly correlated classes with other non-target objects.
   ii. Size of the class object relative to sample.
   iii. Shadows/Reflections.

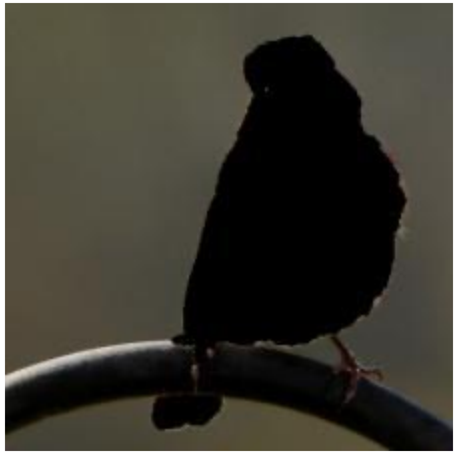
4) APPLICATION ON INDUSTRY QUALITY INSPECTION

5) CONTRIBUTION & DISCUSSION

i. Our Kaggle Dataset: ImageNet-11 (*Public*)

RELATED WORK

- This paper explores how backgrounds might help in object recognition tasks.
- Our work extends on the work done by NOISE OR SIGNAL: THE ROLE OF IMAGE BACKGROUNDS IN OBJECT RECOGNITION by Xiao et al.
- Xiao et al. strengthened the previously established claim that models trained on backgrounds only can help achieve non-trivial accuracy.
- Another important finding, they present in their paper is that more accurate models tend to rely on backgrounds less.



a) Background only

b) Foreground only

RELATED WORK

Xiao et al. created their sub-dataset using the ImageNet dataset. They used 9 classes to understand the role of background, augmented the samples in the following way :

- Only-BG-B: foreground removed with bounding box and replaced with black box
- Only-BG-T: foreground removed with bounding box and replaced with background
- No-FG: foreground segment removed
- Only-FG: background removed
- Mixed-Same: background replaced with same class background
- Mixed-Rand: background replaced with random class background
- Mixed-Next: background replaced with the next class background.

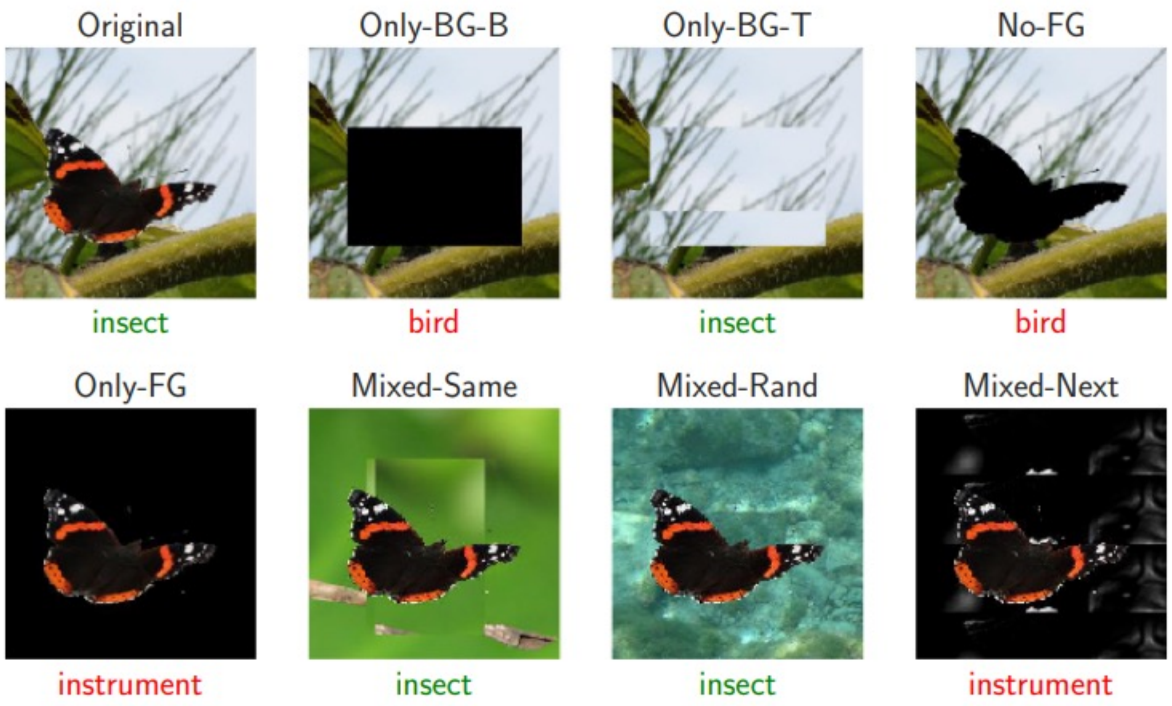


Figure 2.1: **The Illustration of Image9 Dataset in Xiao et al.**

NOISE OR SIGNAL: THE ROLE OF IMAGE BACKGROUNDS IN OBJECT RECOGNITION

The following table summarizes the results reported by Xiao et al.

Trained on	Test Dataset ONLY-BG-B	ONLY-BG-T	ONLY-FG	ORIGINAL	IN-9L
MIXED-NEXT	11.19	8.22	59.60	52.32	46.44
MIXED-RAND	15.33	14.62	74.89	73.23	67.53
MIXED-SAME	35.19	41.58	61.65	75.01	69.21
NO-FG	36.79	42.52	31.48	48.94	47.62
ONLY-BG-B	54.30	42.54	21.38	42.10	41.01
ONLY-BG-T	38.49	50.25	19.19	49.06	47.94
ONLY-FG	23.58	22.59	84.20	54.62	51.50
ORIGINAL	32.94	40.54	63.23	85.95	80.38
IN-9L	34.02	43.60	84.12	96.32	94.61
ImageNet	12.69	17.36	90.17	96.89	95.33
ImageNet (Full-IN)	7.98	9.51	59.19	76.07	-

Table 2.1**: The Illustration of Accuracy Metrics of Image9 Dataset in Xiao et al.**

INTRODUCTION

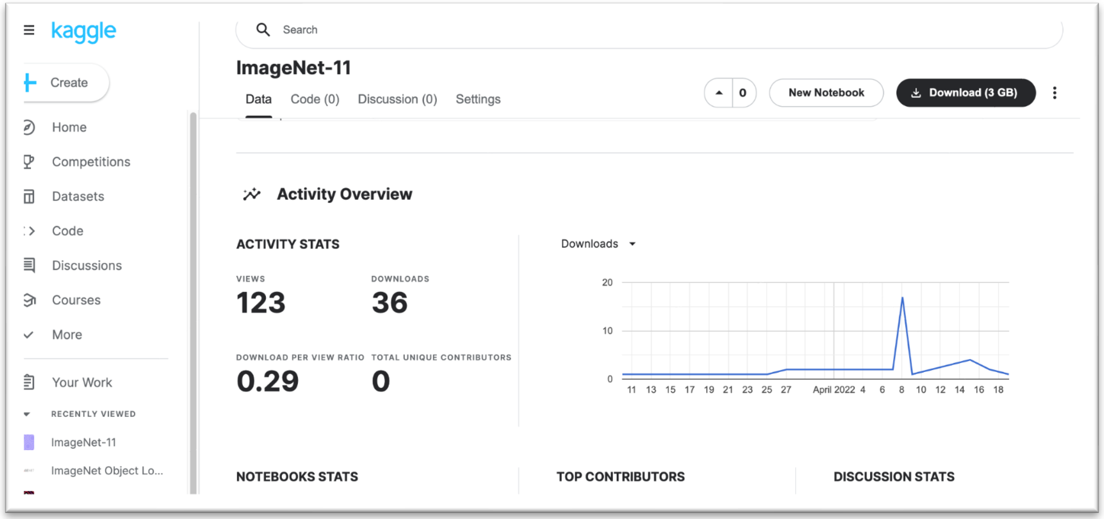
- We categorize the elements present in the background into two main domains:
  - Class independent factors:

    Elements that are not unique to certain classes and contain properties that are generally present in the entire image dataset, such as colors, edges in the background, etc.
  - Class dependent factors:

    Elements that are unique aspects of the class present in the background, e.g., shadows/ reflections, land/sea background, classes often found with other non-target classes, and the class object's size relative to the background.
- In our work we find that models trained and tested with good foreground segmentation perform better than the models trained on original dataset.
- We attribute this to absence of class independent factors.
- Furthermore, we attribute the phenomena of models achieving non-trivial accuracy on backgrounds only to class dependent factors.

DATASET AND SEGMENTATIONS

- Our dataset is available on Kaggle
- We use 13 classes from ImageNet and use the instance masks generated by ***Pixel-ImageNet***.
- We also tested on original dataset used by Xiao et al. in their paper.
- In some topics, we also use industry specific samples to validate our findings.



CLASS INDEPENDENT FACTORS IN BACKGROUND

CLASS INDEPENDENT FACTORS IN BACKGROUND: FOREGROUND ONLY

(a) Foreground only data samples

(b) Validation loss on original (red) and foreground only (green)

(c) Validation accuracy on original(red) and foreground only (green)

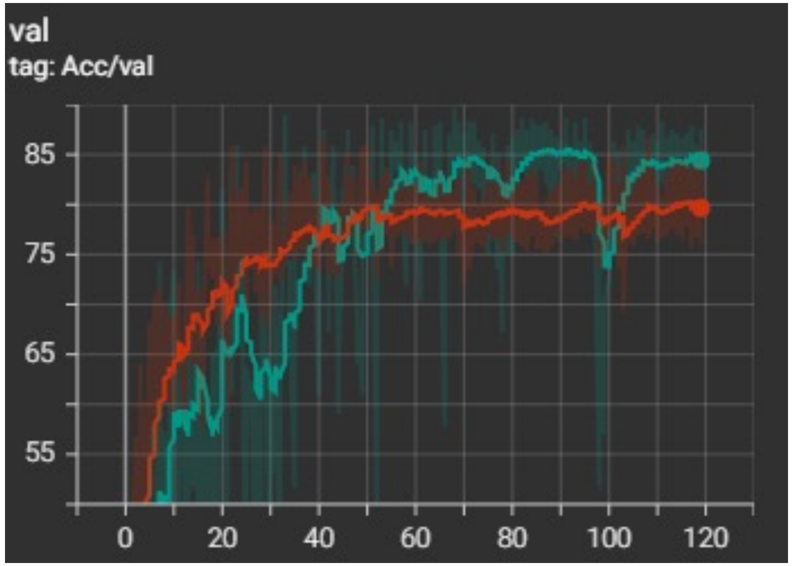


Figure 3.1**: The Illustration of our dataset used for class-independent factors in the background, foreground only.** (a) shows samples of foreground only data, (b) indicates the validation loss and accuracy achieved by the model. We find that datasets with good foreground segmentation provide better classification accuracy compared to the original dataset. We theorize that this is due to the absence of class-independent factors in the samples and therefore lower noise. This allows the model to focus more on the classes. **Our result contradicts the findings of Xiao et al. who showed that models trained and tested on foreground only performed poorly as compared to models trained and tested on the original dataset.**

FOREGROUND SEGMENTATION ISSUES

- In our analysis we found that the discrepancy between our results and Xiao et al. is mainly because of better segmentation masks hence more accurate foreground and background separation.

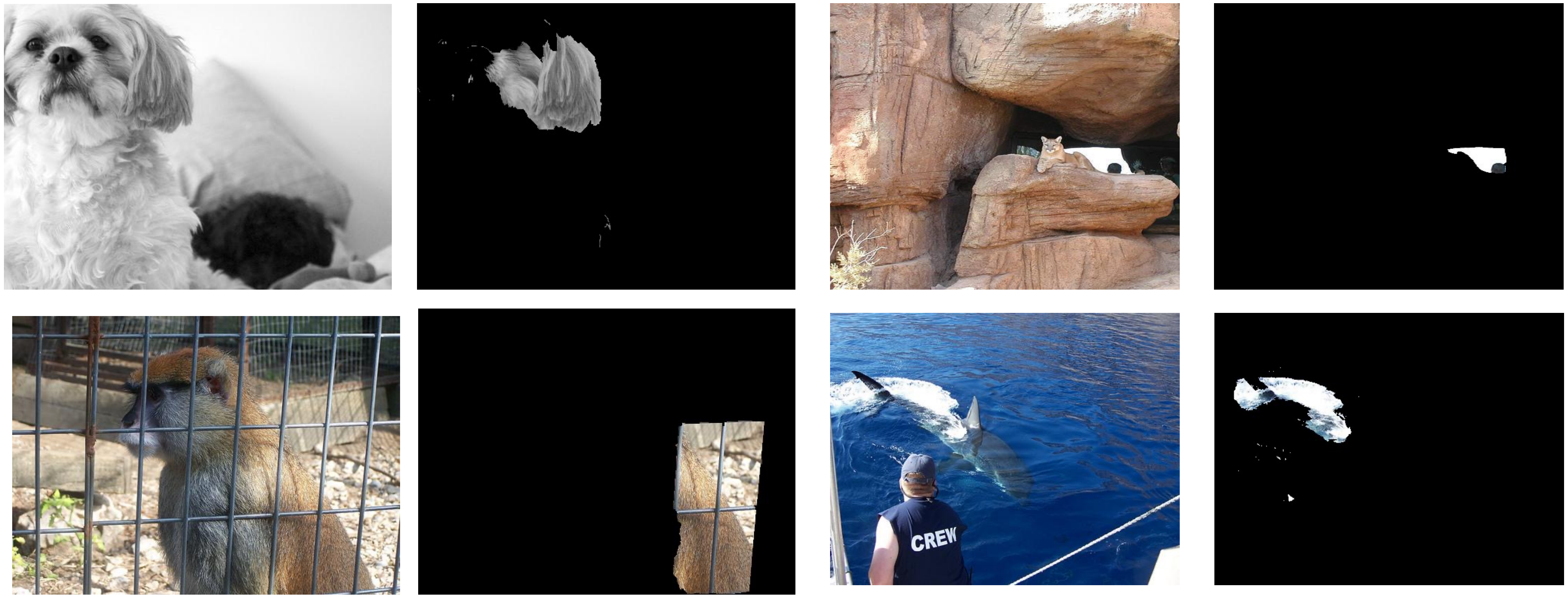
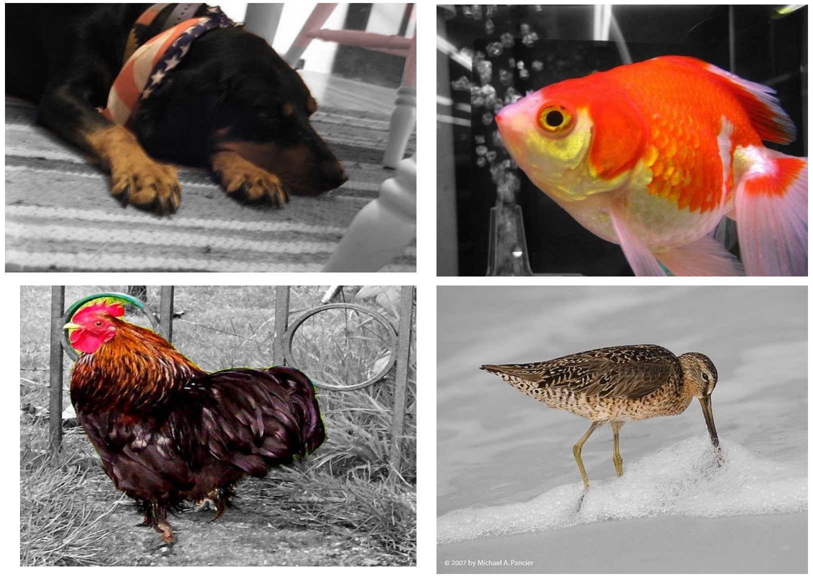


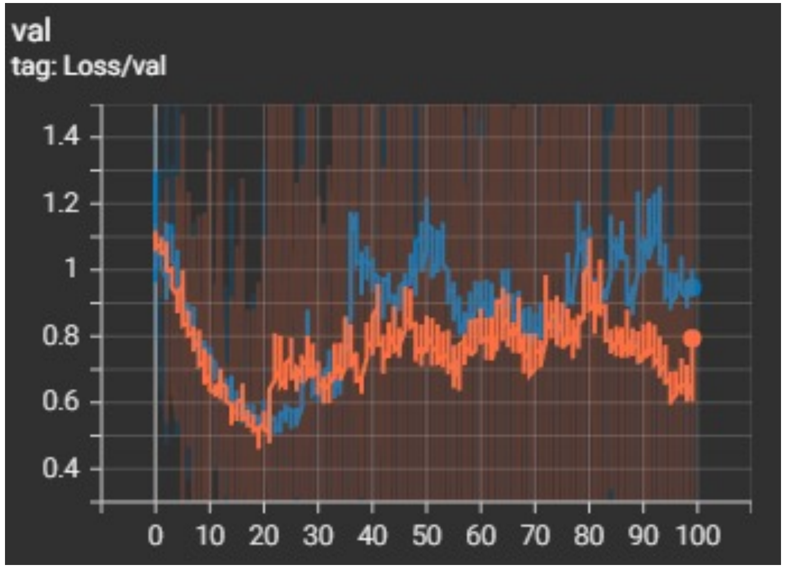
Figure 2.2**: The Illustration of Bad Segmentation Samples Image9 Dataset in Xiao et al.**

CLASS INDEPENDENT FACTORS IN BACKGROUND: Color

(a) Background colored. data samples



(b) Validation loss on original (blue) and background colored (orange).



(c) Validation accuracy on original (blue) and background colored (orange).

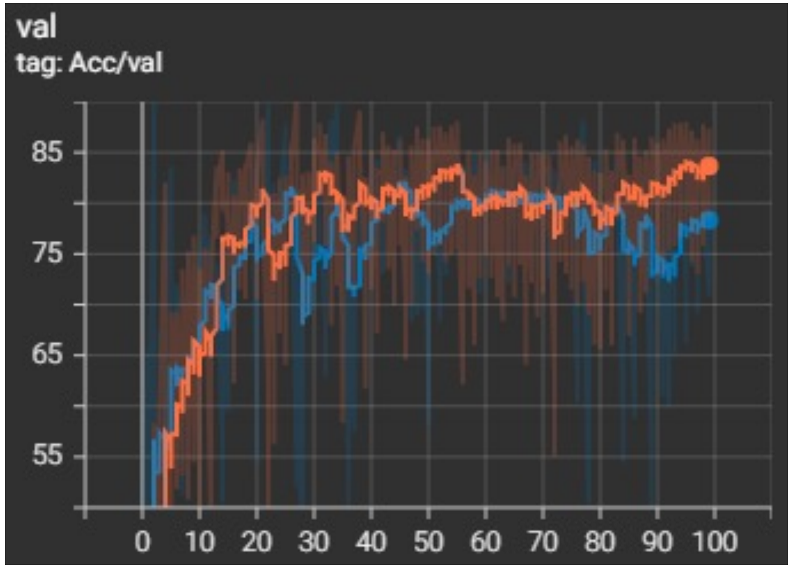
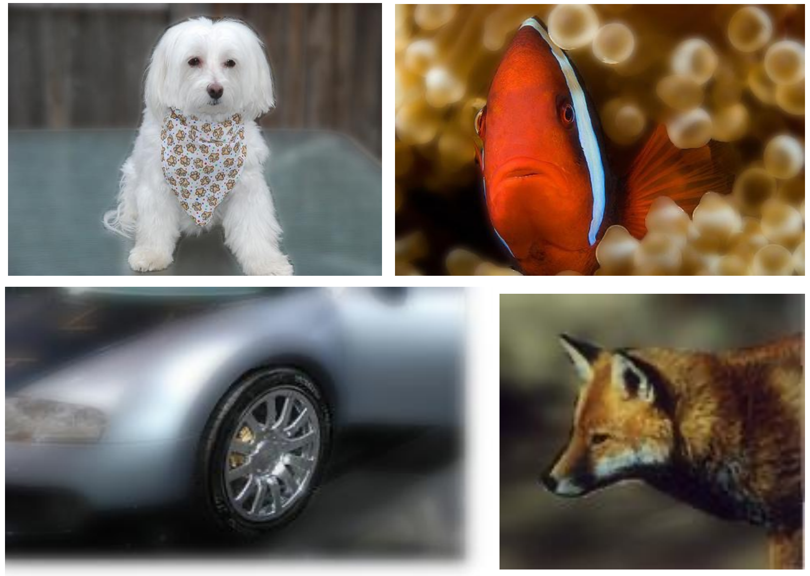


Figure 3.2**: The Illustration of Image-9 dataset used for class-independent factors in the background, Color.** (a) shows the samples of background colored data,(b) the validation loss and accuracy comparison on the original dataset (without any augmentations) and colored background dataset. In this paper, **we find that background colors that produce good edge boundaries with class objects have better classification accuracy compared to the original dataset**. Theoretically, this is because some colors provide better edge boundaries for the target class. The background color matching the target category causes greater difficulty in detection and therefore lower classification accuracy.

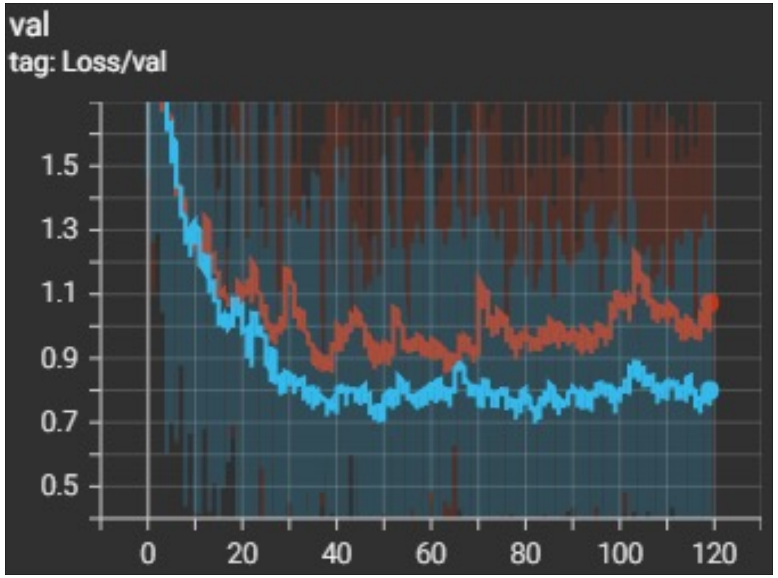
The dataset used for color analysis is a subset of the dataset used by Xiao et al.

CLASS INDEPENDENT FACTORS IN BACKGROUND: Blur

(a) Background blurred data samples



(b) Validation loss on original (orange) and background blurred (blue).



(c) Validation accuracy on original (orange) and background blurred (blue).

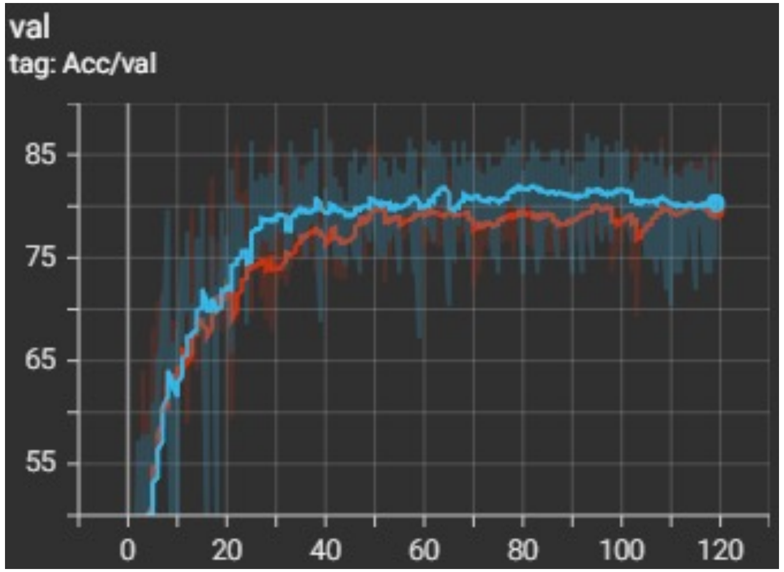
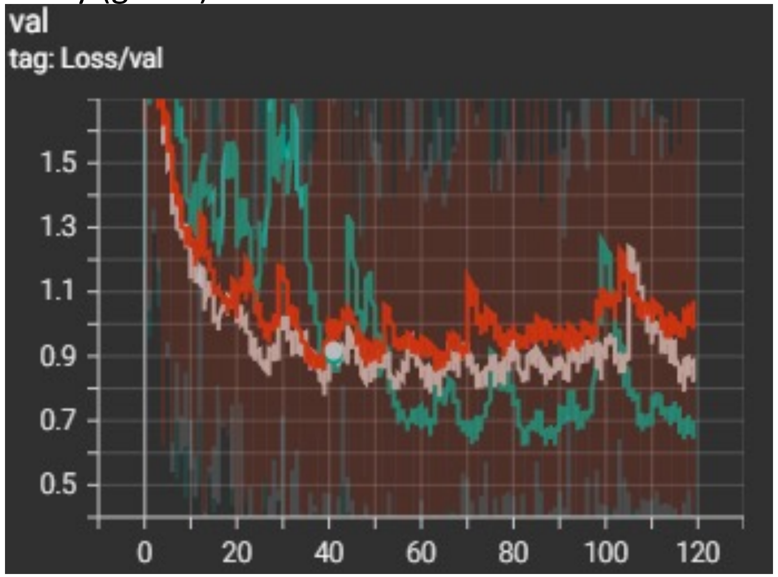


Figure 3.3**: The Illustration of Our dataset used for class-independent factors in the background, blur.** The above figures show the data samples with blurred background and the comparison of validation loss and accuracy on the original dataset (without any enhancement) and the blurred background dataset. **In this paper, we find that the model trained on the dataset with blurred background has better classification accuracy compared to the original dataset. Theoretically, this is because the blurred background offers better edge boundaries and smoothens out the edges in the background while preserving the background color.**

CLASS INDEPENDENT FACTORS IN BACKGROUND: Size of the background

(a) Region reduced background data samples

(b) Validation loss on original(red) and region reduced background(brown) and foreground only (green).



(b) Validation Accuracy on original(red) and region reduced background(brown) and foreground only (green).

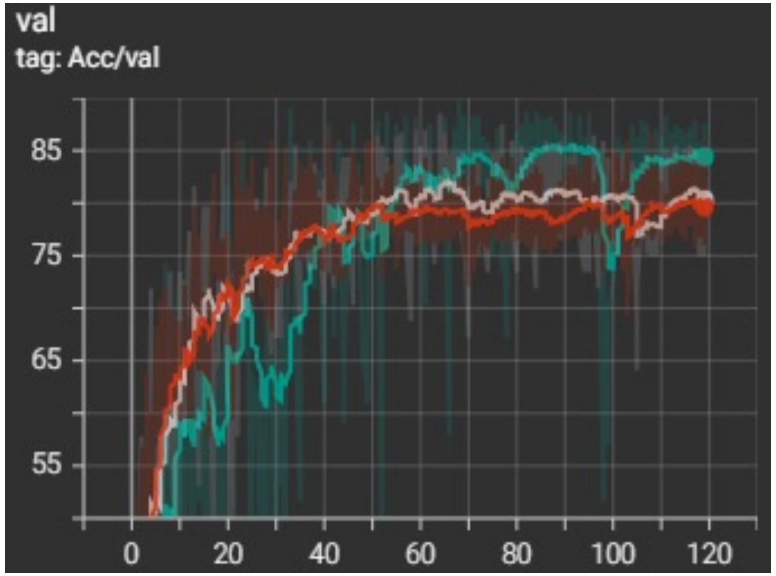


Figure 3.4**: The Illustration of Our dataset used for class-independent factors in the background, size of backrground.** The above figures, (s) shows samples of reduced region of the background data, (b) the comparison of validation loss and accuracy on the original dataset (without any enhancement) and the reduced region of the background dataset. In this paper, we find that models trained on datasets with **reduced background regions have better classification accuracy compared to the original dataset**, and **the accuracy/loss is squeezed between foreground-only (upper bound) and original data (lower bound)**. Instead of increasing the size of the bounding box, we used a **uniformly expanded region** for better analysis.

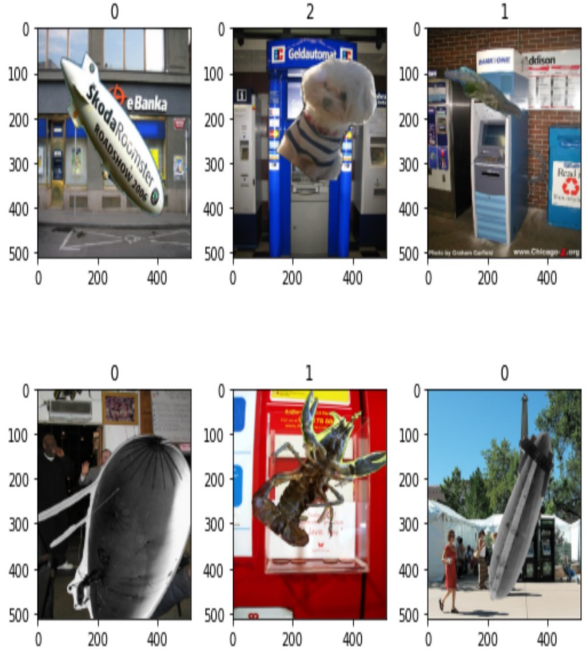
When are backgrounds helpful then?

- There are cases where backgrounds might be essential for good classification accuracy. Some of the cases are mentioned below:
  - If good foreground segmentations are not available, like in xiao et al.
  - If the size of the target instance is small relative to the sample size
  - If the target instance is occluded
  - If the foreground and background separation is not well defined
- In all these cases, backgrounds may be conducive to classification because of the presence of class dependent factors.

CLASS DEPENDENT FACTORS IN BACKGROUND

Class dependent factors in background: non-target objects

(a) The Train dataset



(b) The Test dataset

(b) The composition of each image

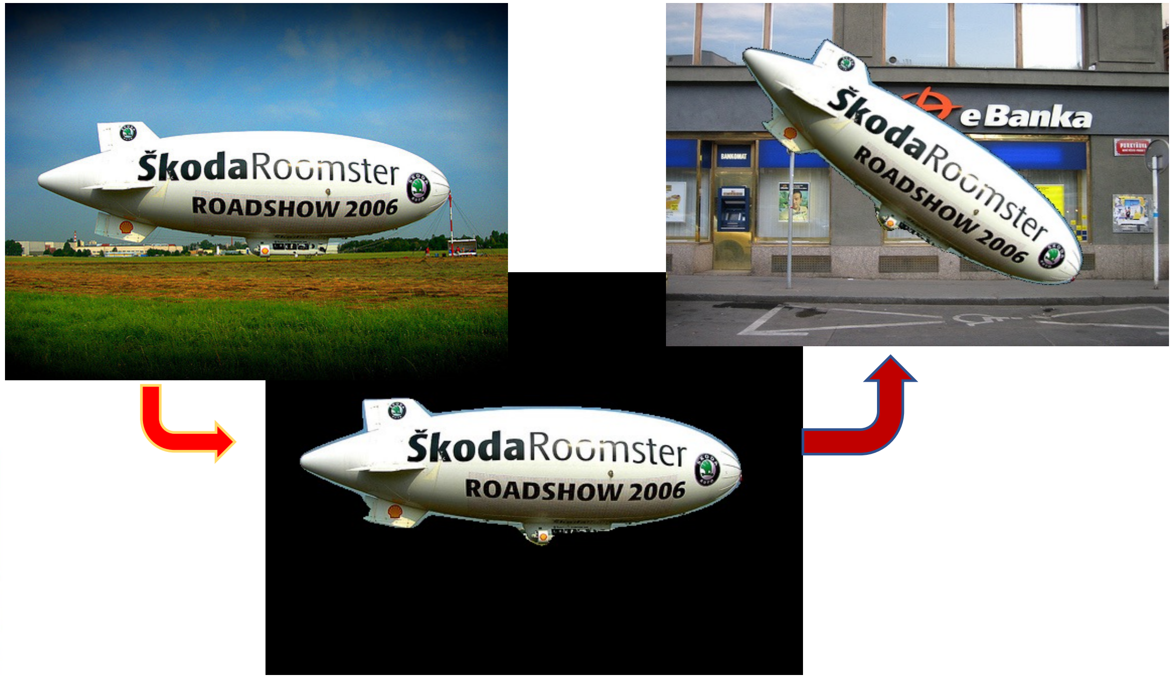


Figure 3.5**: The Illustration of our dataset used for class-independent factors in background, non-target objects.** (a) represents the training dataset, which contains four objects, 0 for the spaceship, 2 for the dog class, 1 for the crayfish, i.e. 3 for the ice cream; (b) a total of 1476 unique images from these four classes were used for training and 371 for testing; (c) for each image, we segmented the ground truth foreground by means of a PixImageNet map, and then we concatenated them with the original generic images of the cashmacnhine class in ImageNet. **The aim of this process is to remove the original background correlation present in these four classes, which indicates for instance that in the original natural environment, caryfish could never been present in the background of the cashmachine**.

Class dependent factors in background: non-target objects

Figure 3.6**: Diagrammatic representation of the intra-class diversity applied in the dataset.** (a) Each class has one or more different objects, which may have a different shape and colour. Each category has about 500 segmented objects combined with the generated background, and in total we segmented almost 1800 images in this experiment. **As shown in the figure, the dataset contains ice screams of different shapes, as do the other dog, crayfish and spaceship classes. Here, we avoided the problem of unbalanced data allocation, i.e. for each category we used approximately the same amount of data for training.**

Class dependent factors in background: non-target objects

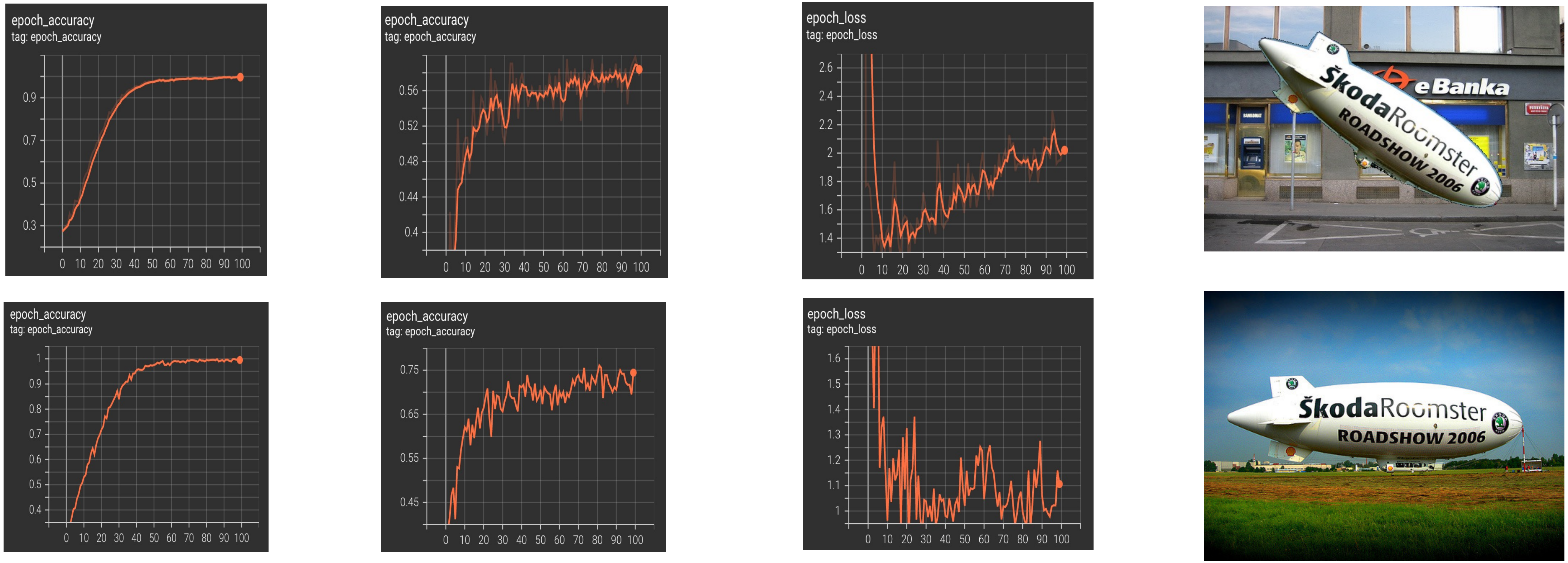


Figure 3.7**: Illustration of training/validation accuracy and loss.** (a) The upper raw data represents the results for combination with the generalized background. (b) The second raw plot represents the results in the original images without breaking the correlation between foreground and background; (c) According to the training accuracy plot, there is no huge difference in validation, but we can note a 20-25% drop in accuracy and a significant increase in loss after removing the original background and combining it with the cash machine background, implying that in the current experimental setting**, the correlation of the generalised background has a significant impact on the model (EfficentB0) performance.**

Class dependent factors in background: non-target objects

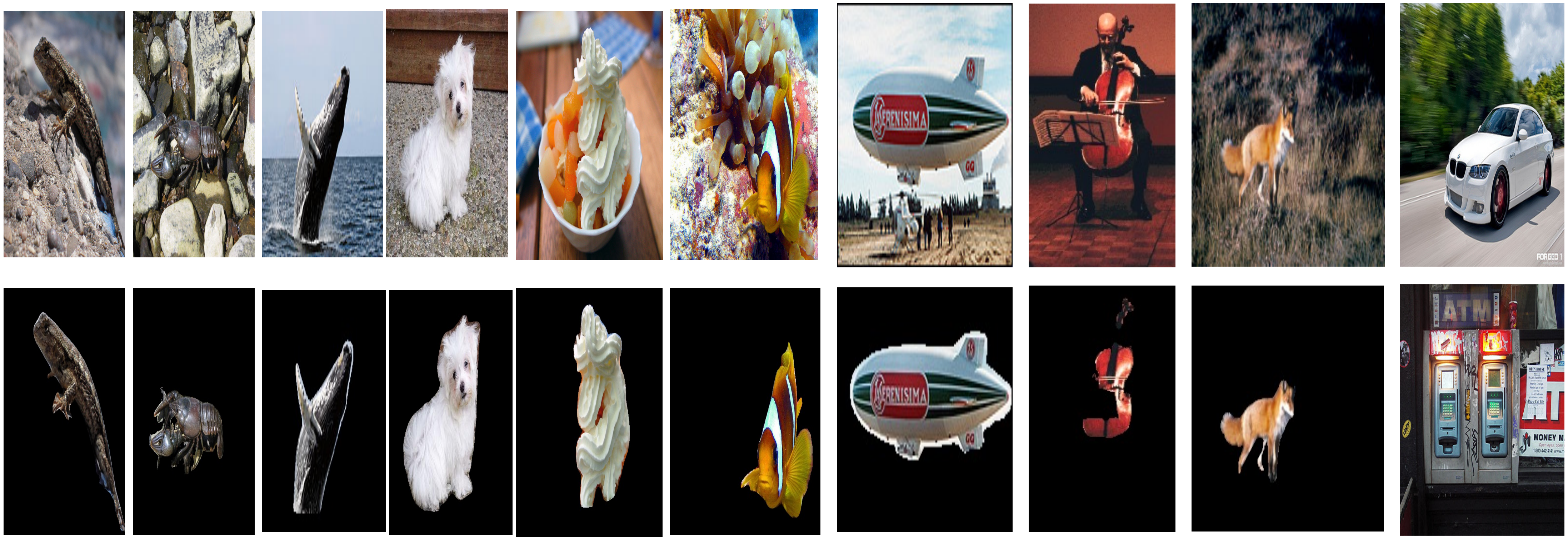


Figure 3.8**: Schematic of the 10 multi-class dataset in the training experiments for segmenting foreground objects.** (a) The first raw data represents the base fact images in ImageNet, here we selected nearly **11 classes** from ImageNet, from ice cream to whales, from cars to instruments, each with approximately **1300 images, for a total of 14,3K images** used in training/testing, (b) For comparison, we used PixelImageNet map to segment out the foreground object from original image and train the same number of epochs with the same hpyeparameters. the whole training took 3 h 31 min 8 s.

Class dependent factors in background: non-target objects

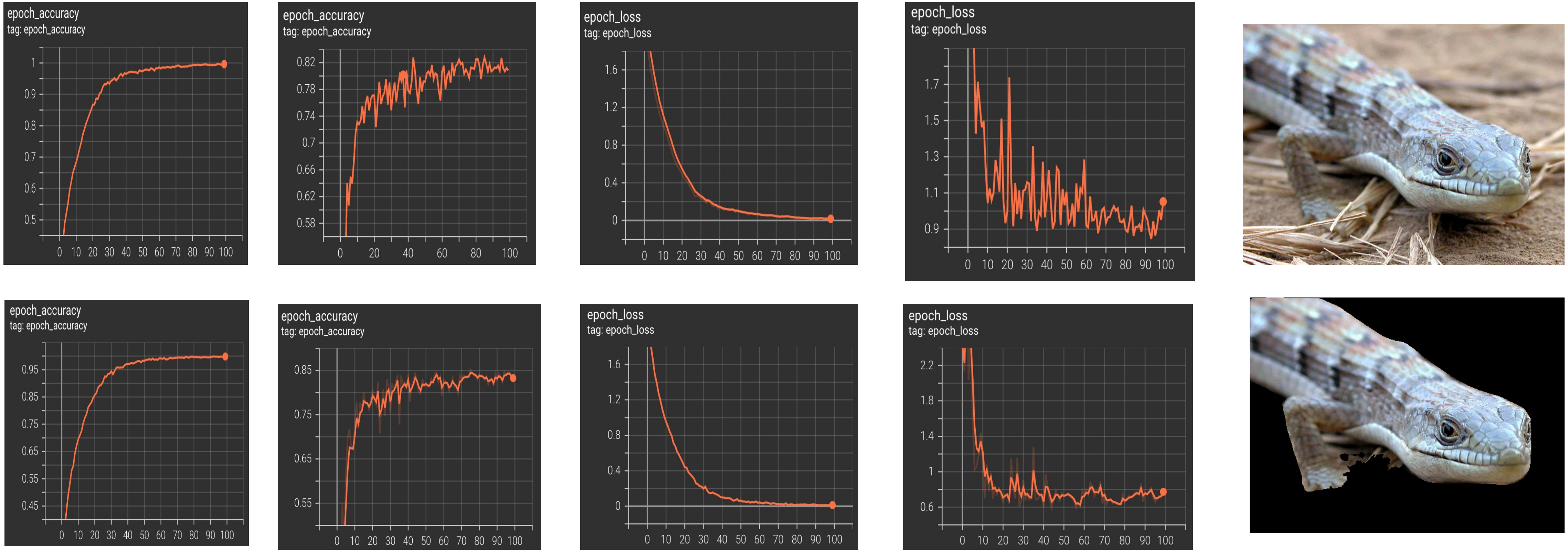


Figure 3.9**: Comparison between original image training and well-segmented ground truth training.** (a) The first raw data represents the metrics of the original image training and the second raw data shows the results of the well-segmented ground truth training. **(b) Compared to the segmented objects in the Noise paper, with a better segmentation mask, we verify the opposite result, i.e. that the model achieves better results, i.e. a 3% improvement in accuracy and a more consistent reduction in loss.** (c) **For large, distinguishable features, a promising segmentation leads to a slight improvement.**

Class dependent factors in background: non-target objects

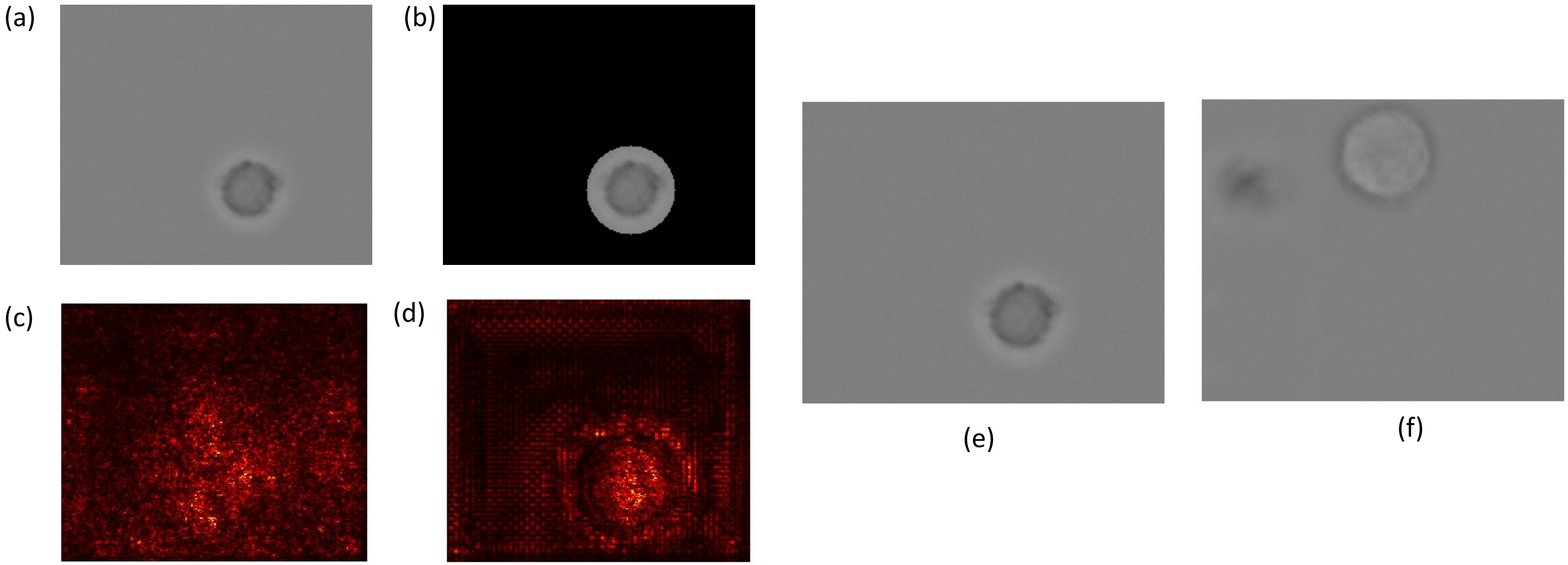
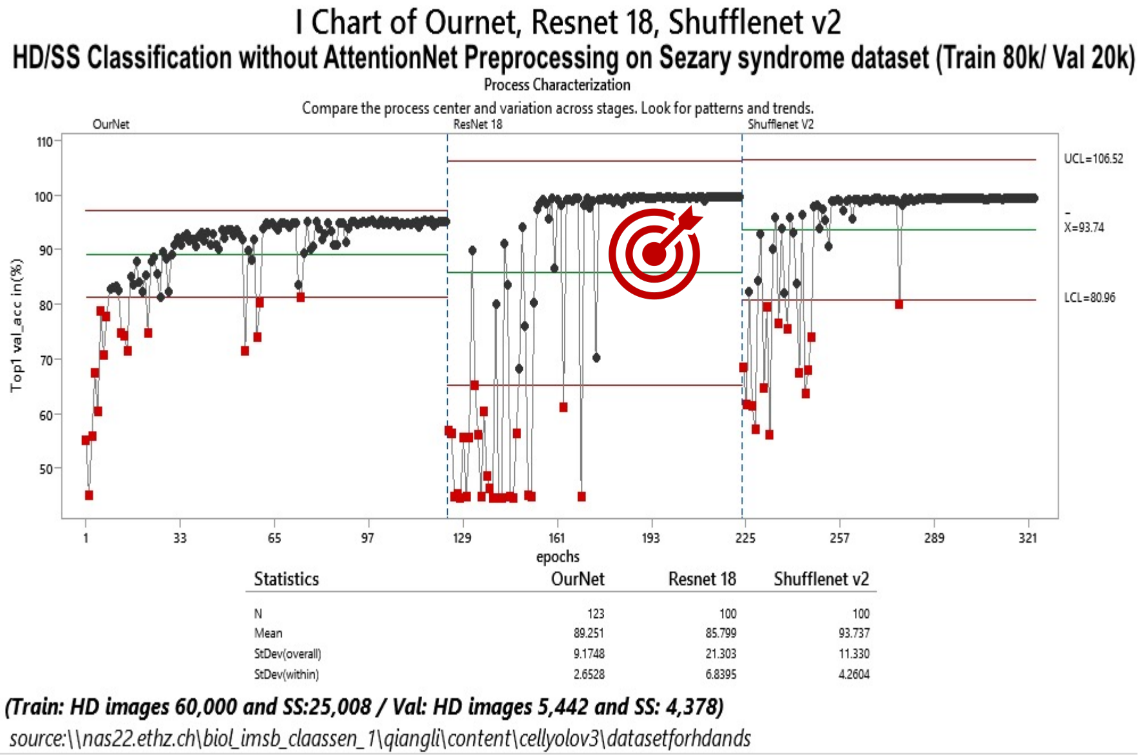


Figure 3.10**: The Comparison between raw image training and ground truth training on the Sezary Sysdrom dataset with good segmentation.** (a) Cells represent high-definition healthy cells from the sezary sysdrom dataset. (b) Healthy cells with background removed. (c) Saliency map of HD cells using ResNet-18. (d) Saliency map of segmented HD cells using ResNet-18. (e) Healthy cells from the sezary sysdrom dataset. (f) Unhealthy blood T cell from the sezary sysdrom dataset, **here the morphological differences between the two types of cells are shown, and both are small features.**

Image source: Qiang et al. AttentionNet on Sezary Sysdrom, AI4PH workshop ICLR 2021.

Class dependent factors in background: non-target objects



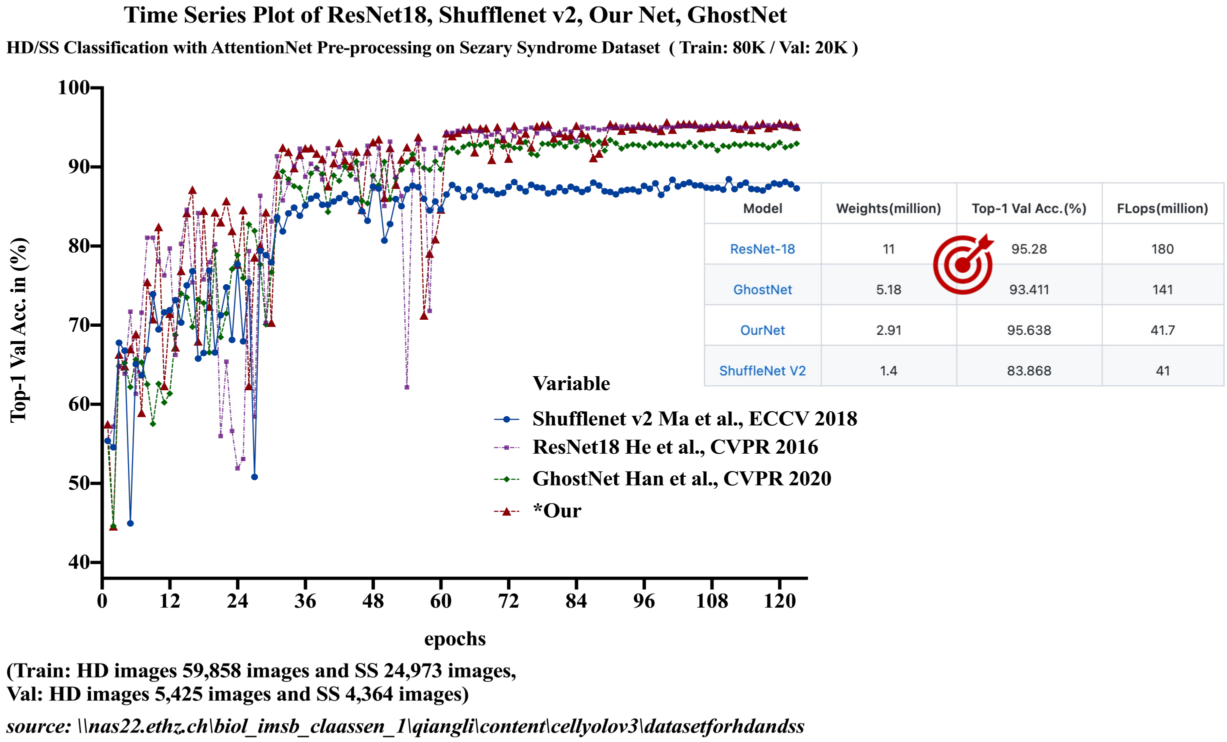


Figure 3.11**: Comparison of time series plots and graphs on the sezary sydrom dataset with/without removing the full background.** (a) The left panel shows CellYolo, RestNet, ShufflnetV2 on the raw dataset of the sezary sydrom dataset, where the three networks reach a high accuracy close to 1.0, in particular ResNet and ShufflNet. Comparing to the right plot after background elimination, here, the Top-1. Acc. shows a slight decreasement, nearly 5%. (b**) those SOTA networks rely heavily on the background, especially for small features or objects.**

Image source: Qiang et al. AttentionNet on Sezary Sysdrom, AI4PH workshop ICLR 2021.

Class dependent factors in background: Highly Correlated Class

(a) The Cray Fish Class (b) The Dog Class

(c) The Whale Class (d) The Icecream Class (c) Additional labelled class Sea for Whale

Figure 3.12**: Illustration of Our dataset applied to a highly correlated class of experiments.** (a) represents the various crayfish objects in our dataset, which contains approximately 500 hand-selected images from ImageNet. (b) represents the dog class in our dataset, which contains approximately 500 hand-picked images from ImageNet. (c) denotes the whale in our dataset, and the same number applies to the ice-cream class in (d). (e) Here we have labelled each whale image with a highly correlated class (sea). **The ground truth boxes for the whale often overlap with the anchor boxes for the sea category, which leads to an less precise classification**.

Class dependent factors in background: Highly Correlated Class

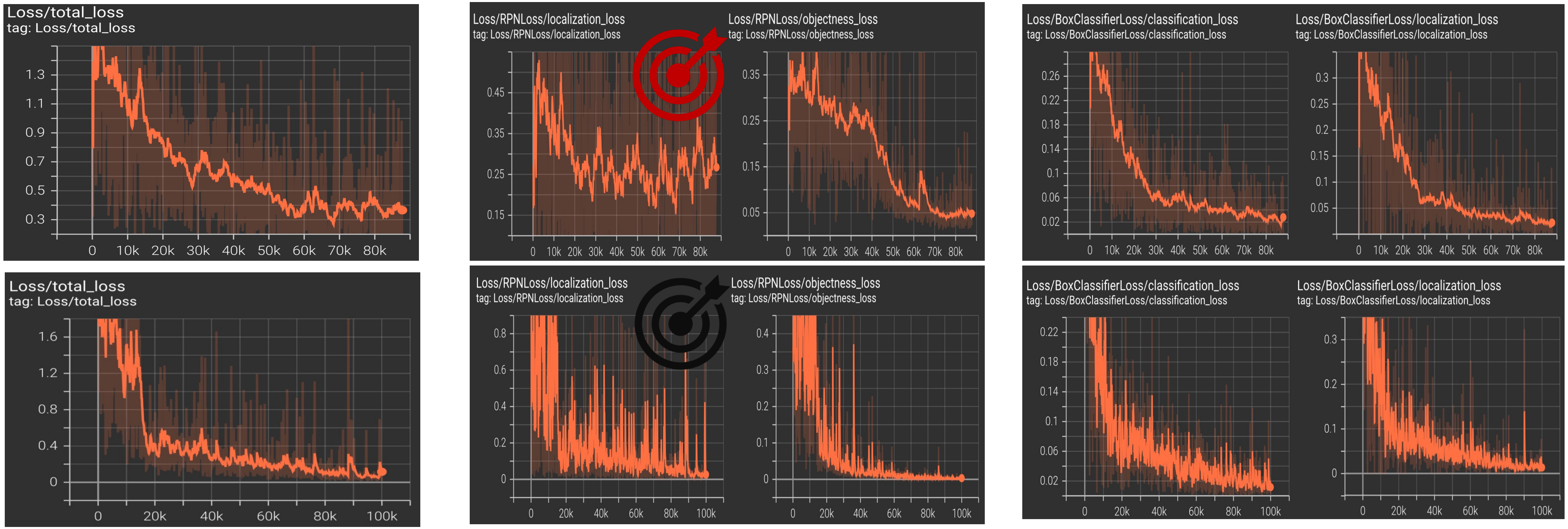
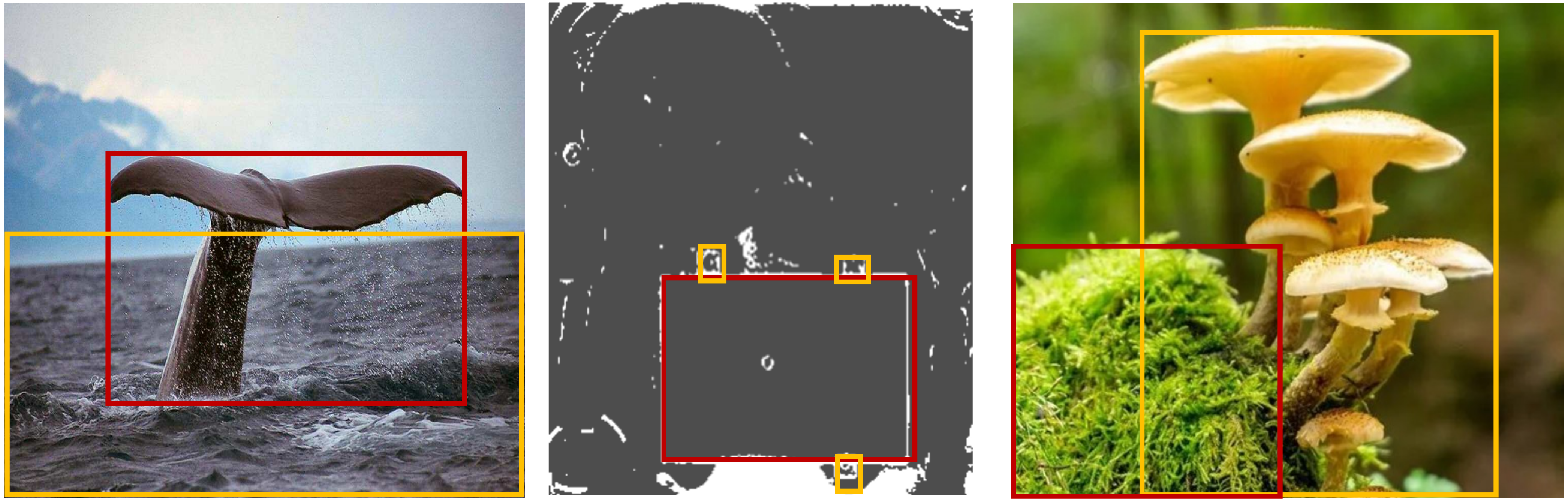


Figure 3.13**: Comparison of training with and without highly correlated classes.** (a) The first raw data represents the metrics of the original image training, which contains only four classes, and the second raw data shows the results of additional high correlation (sea) annotations. (b) We illustrate that there are small differences in the total loss, which is the combination of RPN loss (Localization Loss or the Loss of the Bounding Box regressor for the RPN), Objectness loss (Loss of the Classifier that classifies if a bounding box is an object of interest or background), Box classifier loss (Loss for the classification of detected objects into various classes), and box localization loss (Localization Loss or the Loss of the Bounding Box regressor). (c)) It is worth noting that there is a significant difference in the loss of RPN for position with and without additional sea class annotation, approaching a loss of 0.25.

Class independent factors in background: Highly Correlated Class



(a) (b) (c)

Figure 3.14**: The Illustration of RPN loss under different conditions for highly correlated classes.** (a) and (c) indicate overlapping classes between the target object and the highly correlated class. The industrial example in (b) shows a more precise example of a highly relevant class for the target object. With the class annotations of the gate control unit, the three target screw classes can be detected more reliably and accurately with an accuracy close to 99.97% and a false positive rate of 0.

Class dependent factors in background: Region

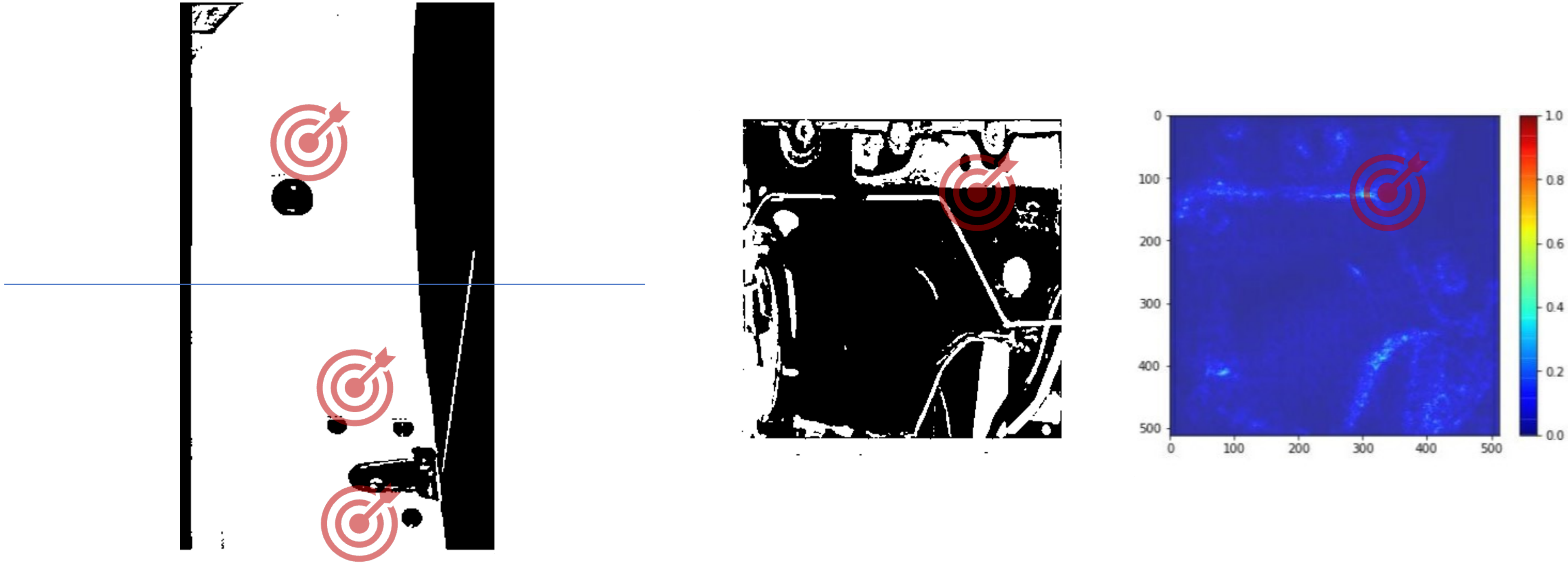


Figure 3.15**: The Illustration of region effects on object recognition.** (a) The inference image on the left represents a negative example of region effects on model prediction, where only the top plug is the target object, where supervised learning was applied and the model was trained on FRCNN with a dataset of over thousands of expert labels and 100,000 epochs. (b) The saliency map on the right shows a positive example of regional information being used to classify different variants, where the highlighted region is the transitive region and the most remarkable feature.

Class dependent factors in background: Shadows/Reflections

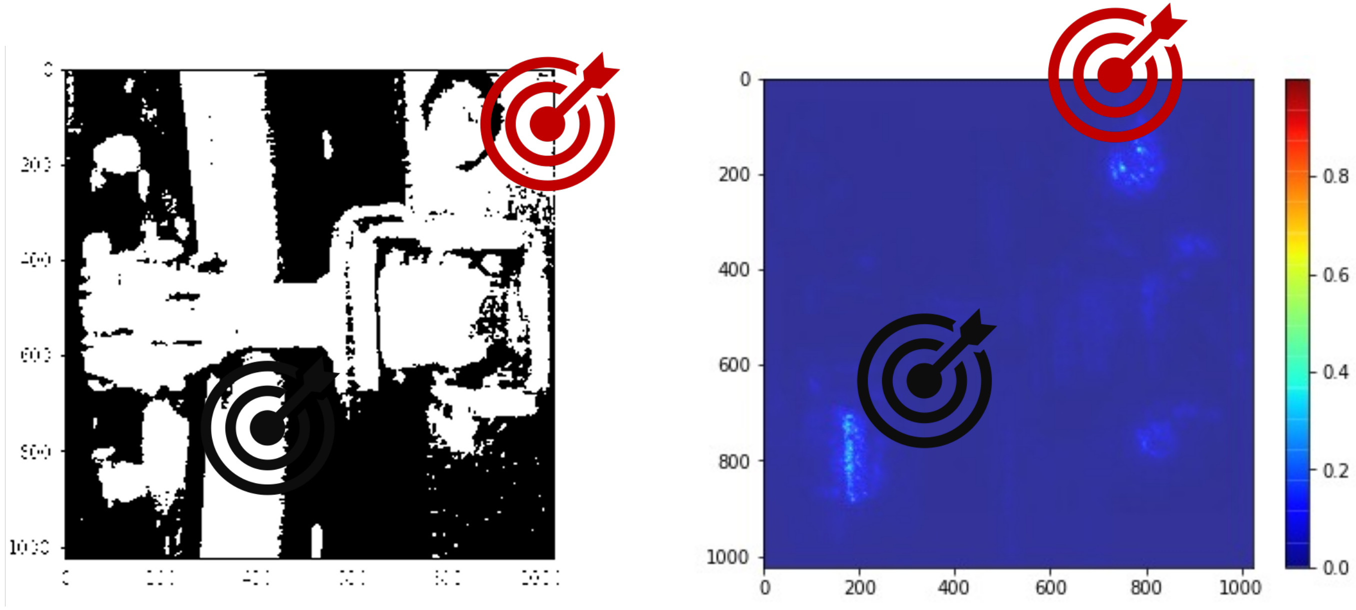
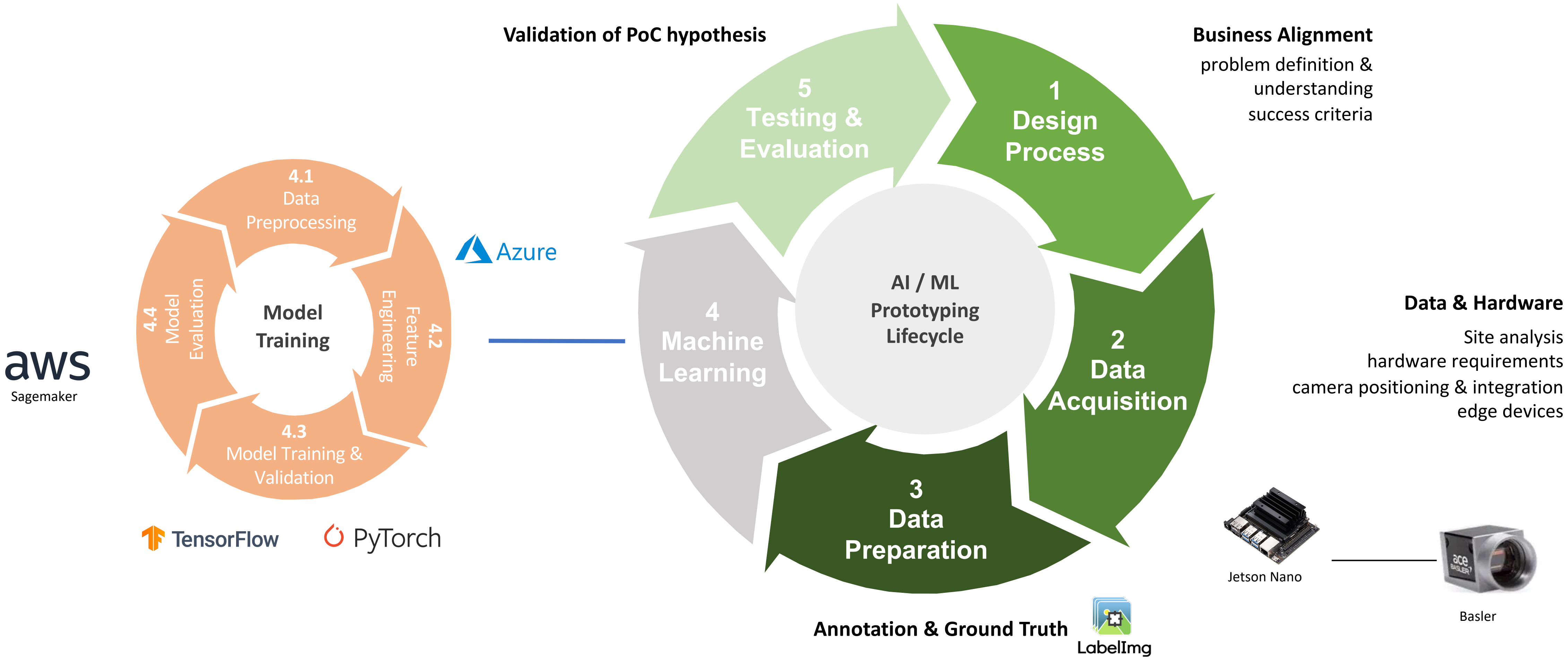


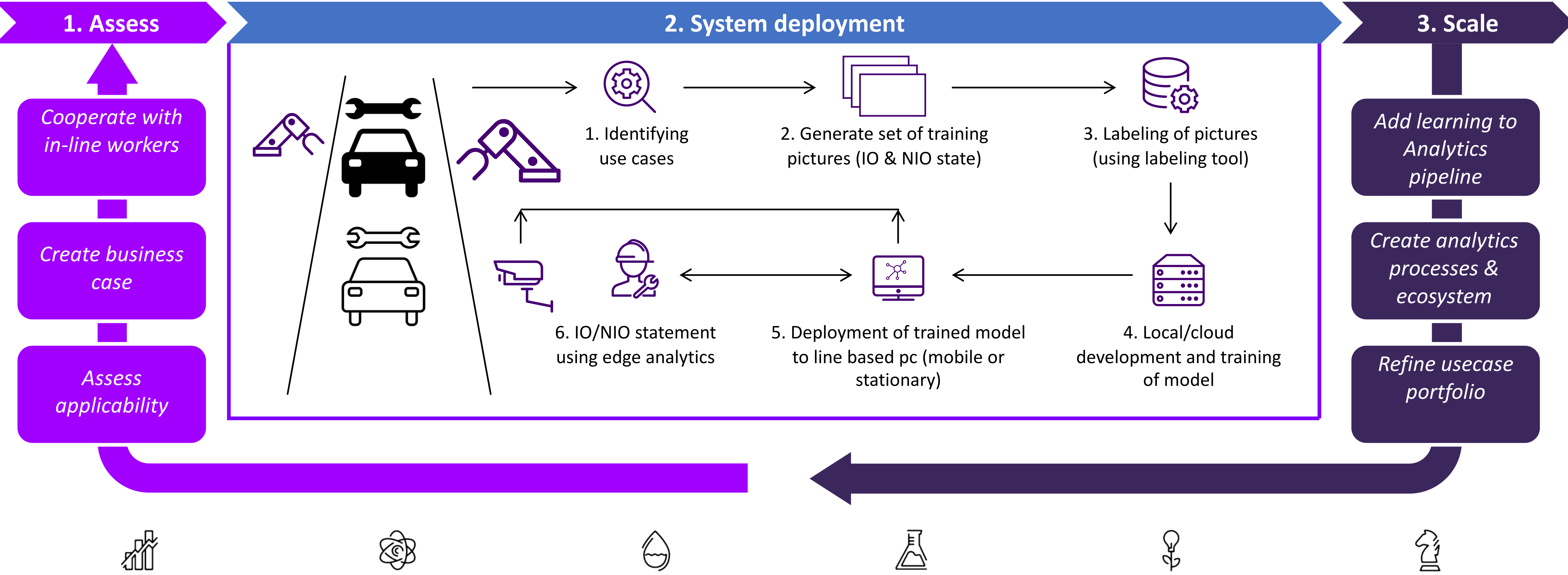
Figure 3.16**: The Illustration of the effect of reflection on object recognition.** Here we show a negative example of reflective information being used for classification, here the highlighted area is the reflective region which reduces the accuracy of the classification and leads to unacceptable cases of false OK in industrial applications.

Roadmap for Computer Vision PoC in Quality Inspection.

Indicative



DATA RIVEN QUALITY PROJECTS WITH A UNIFIED FRAMEWORK AND INNOVATION ECOSYSTEM.



DISCUSSION.

THANK YOU FOR YOUR ATTENTION!

ANY IMPORTANT TOPICS LEFT/ UNTOUCHED?

PLEASE FEEL FREE TO CONTACT US.

ANY QUESTIONS?

APPENDIX