

ZOD : Zero-shot and Out-of-Distribution Detection Dataset for Document Images

Supplementary Material

1. Augmentation

We have also benchmarked ZOD by using certain augmentation techniques to align our sample images with the characteristics of the RVL-CDIP corpus. These augmentations were performed using the Augraphy Python library (<https://github.com/sparkfish/augraphy>), which is specifically designed to produce document images that replicate different kinds of photocopies and printed documents. These augmentations introduce realistic variations and noise patterns commonly found in photocopied documents as we observe in RVLCDIP. We also performed cropping to focus on the main areas of the image.

2. Additional Statistics

2.1. Confusion Matrix for ZOD -I

Figure 1- Figure 6 presents the confusion matrix for ZOD-I subset, providing a detailed breakdown of classification performance across the five overlapping categories: advertisement, file folder, invoice, presentation, and resume, for all the chosen models.

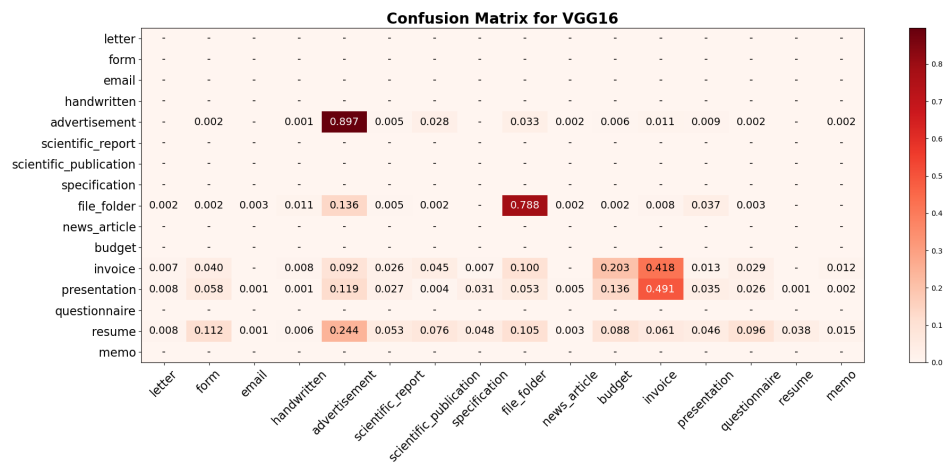


Figure 1. Confusion matrix for the VGG-16 model on the ZOD-I split

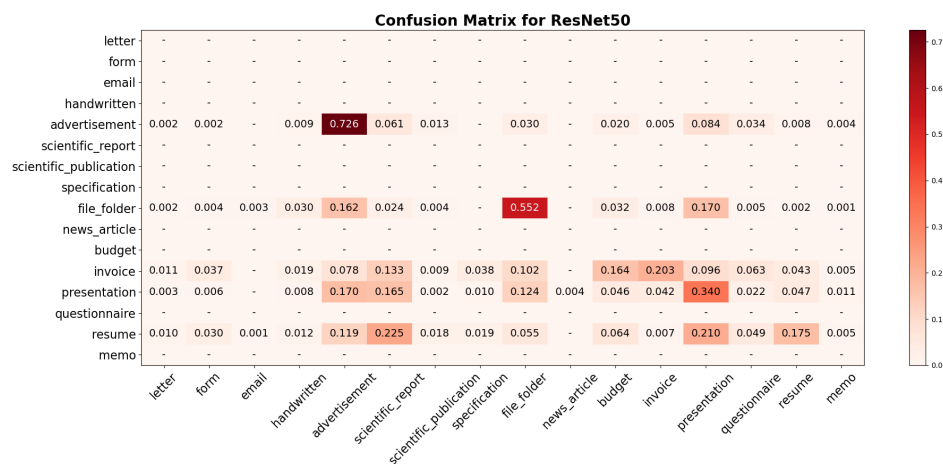


Figure 2. Confusion matrix for the ResNet-50 model on the ZOD-I split

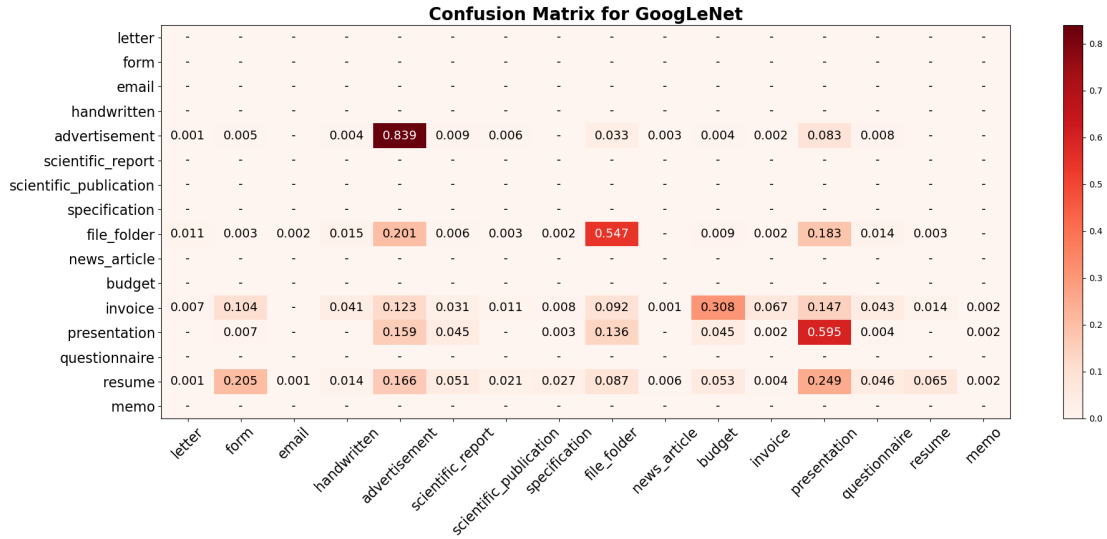


Figure 3. Confusion matrix for the GoogLeNet model on the ZOD-I split

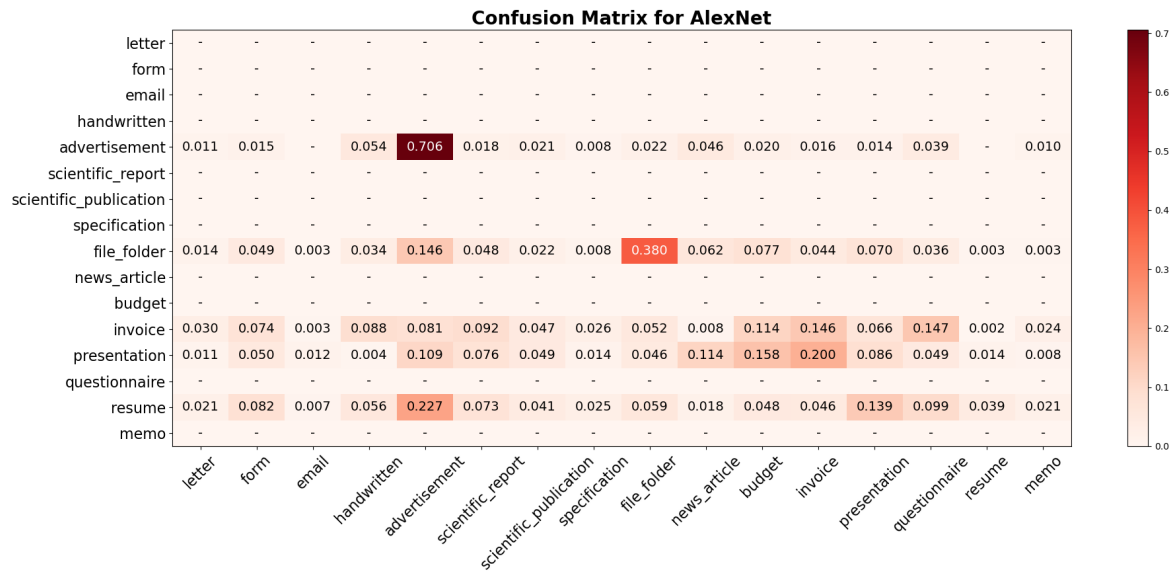


Figure 4. Confusion matrix for the AlexNet model on the ZOD-I split

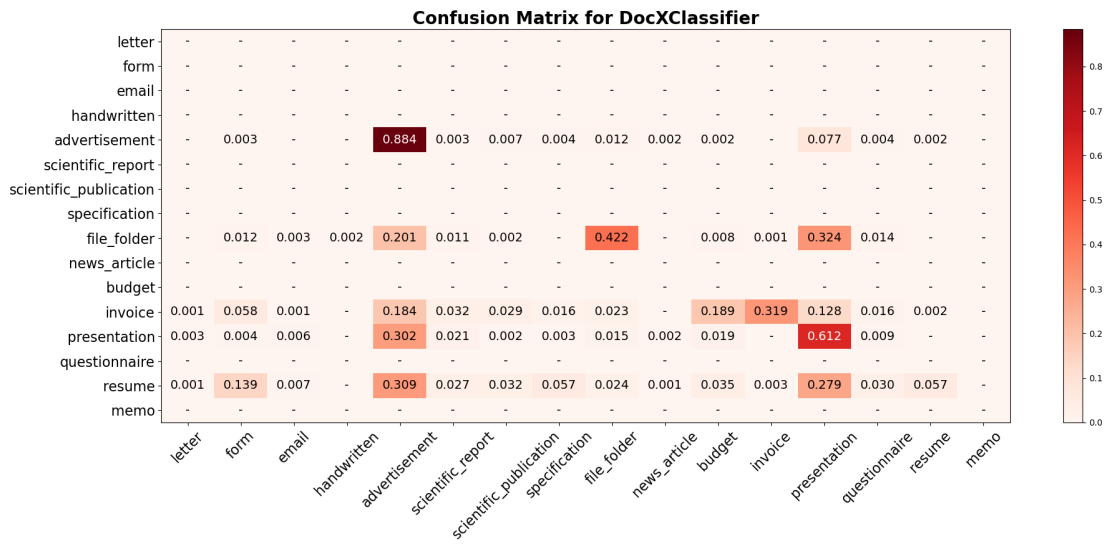


Figure 5. Confusion matrix for the DocXClassifier-base model on the ZOD-I split

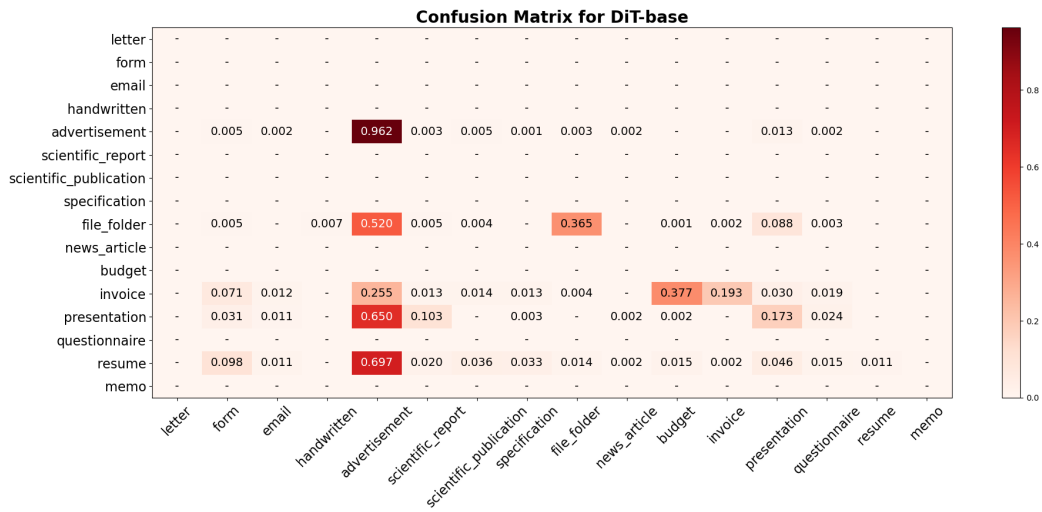


Figure 6. Confusion matrix for the DiT-base model on the ZOD-I split

2.2. Accuracy scores and density distribution on the basis of confidence and energy

Figure: 7 - Figure: 18 each show two different plots. Firstly, the left sub-figure (a) shows the variation in accuracy of the respective model on RVL-CDIP-testset, ZOD-I and ZOD-O, with respect to the range of confidence and energy thresholds. For RVL-CDIP-testset and ZOD-I, containing in-domain class samples, the accuracy decreases with the increase in threshold because the higher the confidence threshold, the

more difficult it is for the model to make correct predictions. Contrarily, as the threshold increases for ZOD-O, the accuracy also increases. This is because more out-of-domain samples fall below the threshold, leading to them being correctly identified as out-of-domain. The sub-figure on the right (b) shows the density of confidence and energy scores for ZOD-I and ZOD-O. This plot shows the distribution of scores across the in-domain and out-of-domain samples of ZOD and the overlap between them.

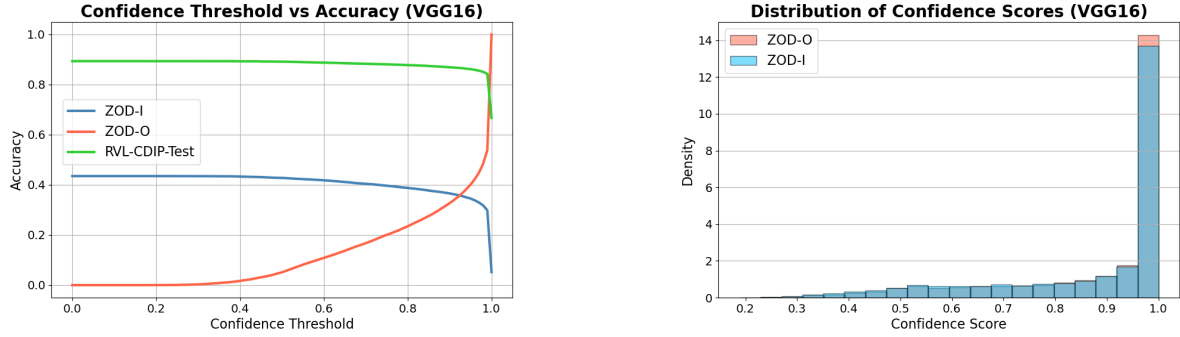


Figure 7. Accuracy scores and density distribution on the basis of confidence for VGG-16

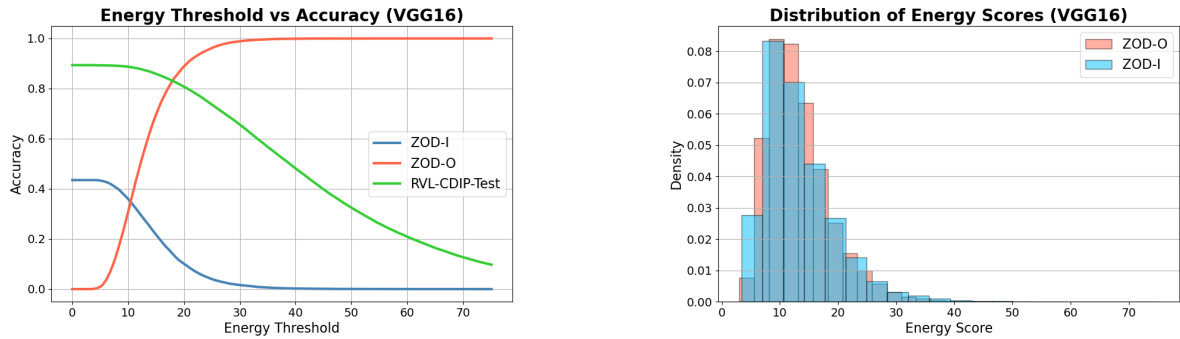


Figure 8. Accuracy scores and density distribution on the basis of energy for VGG-16

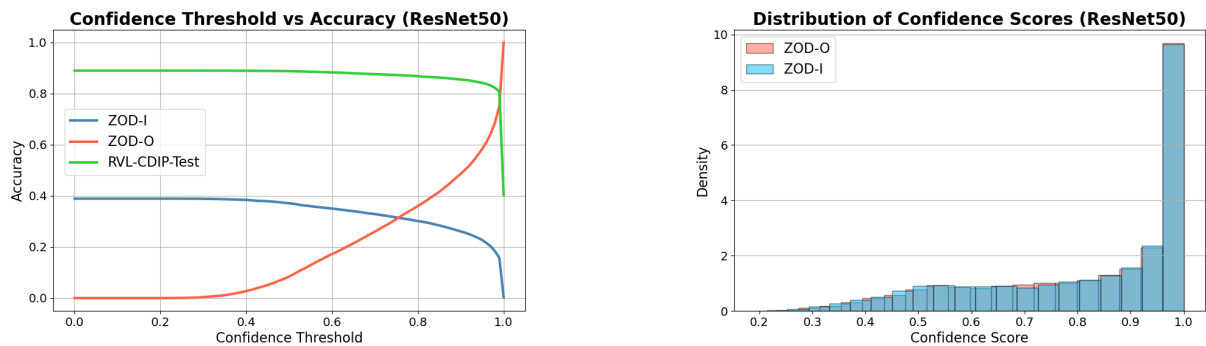


Figure 9. Accuracy scores and density distribution on the basis of confidence for ResNet-50

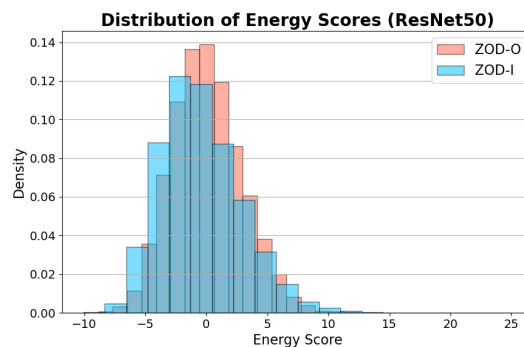
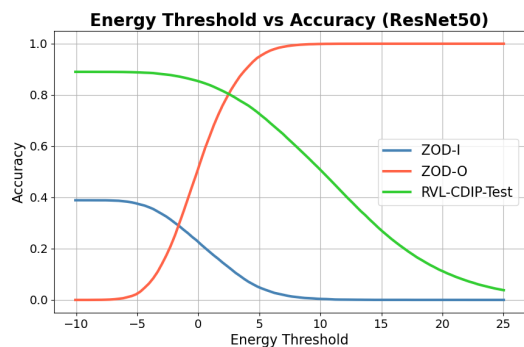


Figure 10. Accuracy scores and density distribution on the basis of energy for ResNet-50

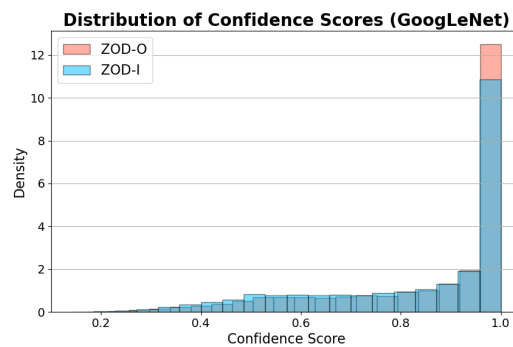
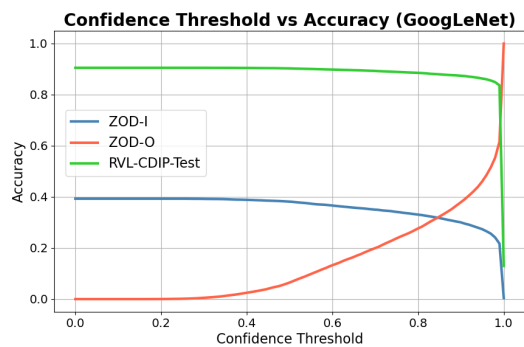


Figure 11. Accuracy scores and density distribution on the basis of confidence for GoogLeNet

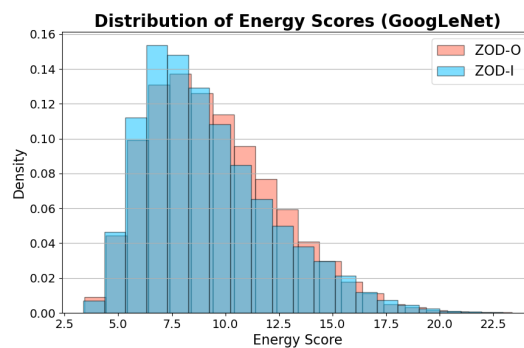
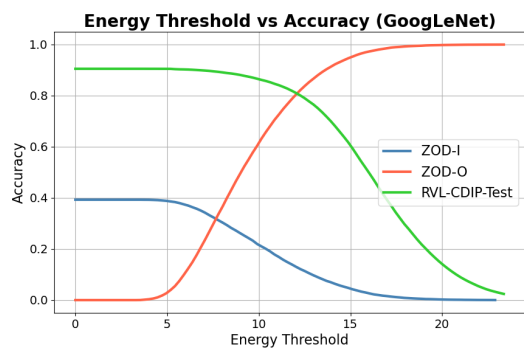


Figure 12. Accuracy scores and density distribution on the basis of energy for GoogLeNet

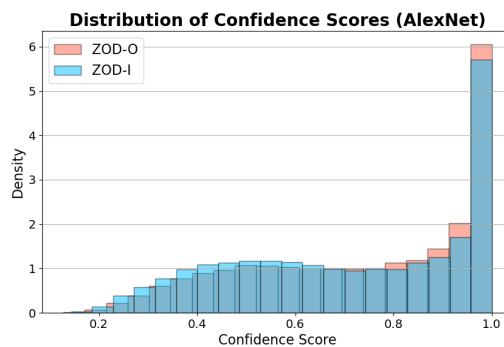
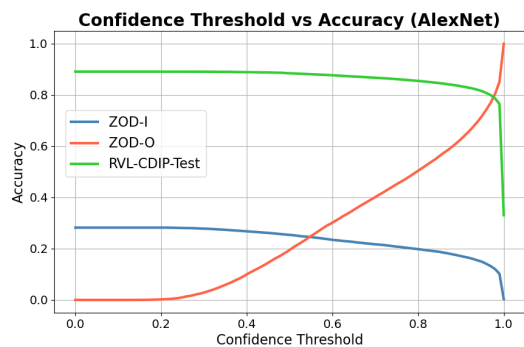


Figure 13. Accuracy scores and density distribution on the basis of confidence for AlexNet

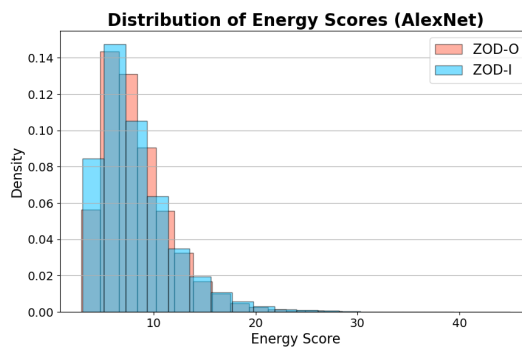
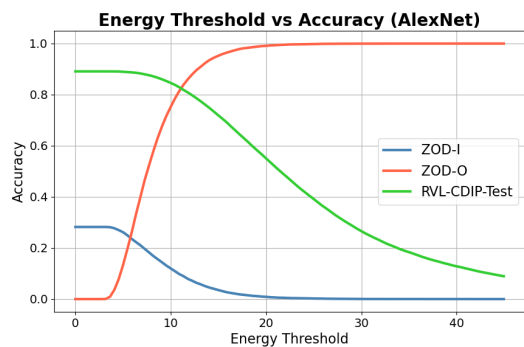


Figure 14. Accuracy scores and density distribution on the basis of energy for AlexNet

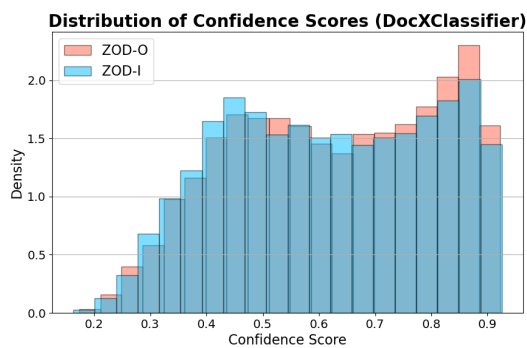
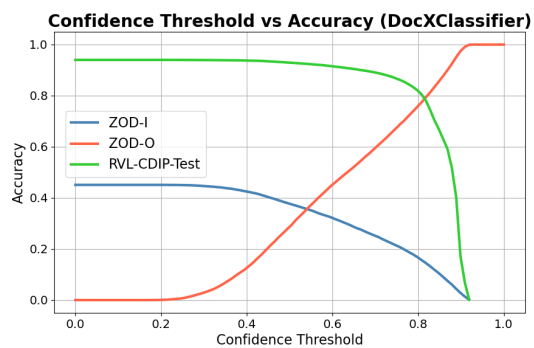


Figure 15. Accuracy scores and density distribution on the basis of confidence for DocXClassifier-base

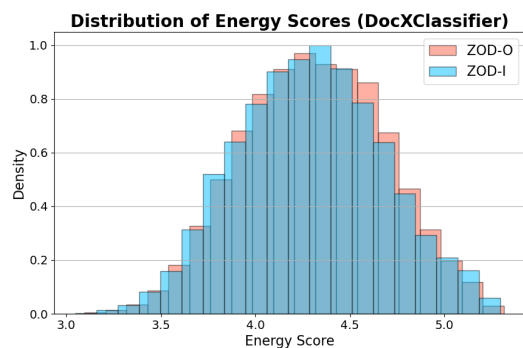
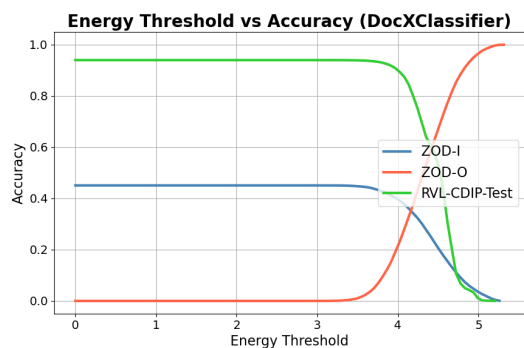


Figure 16. Accuracy scores and density distribution on the basis of energy for DocXClassifier-base

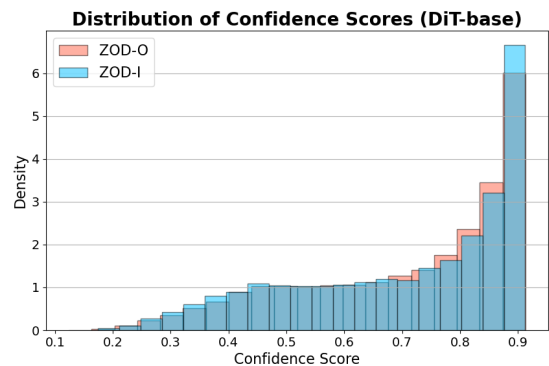
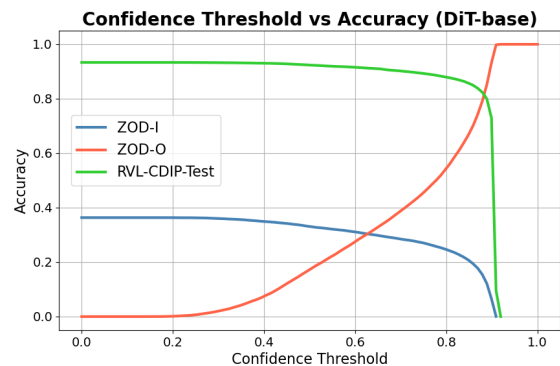


Figure 17. Accuracy scores and density distribution on the basis of confidence for DiT-base

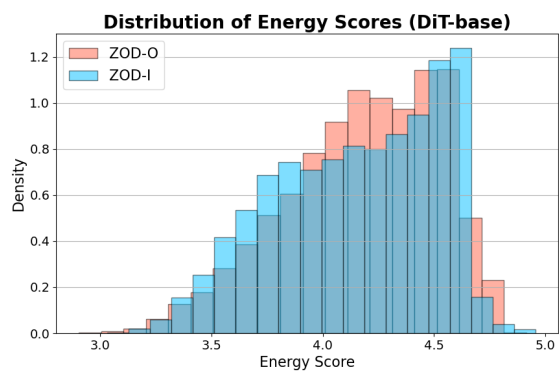
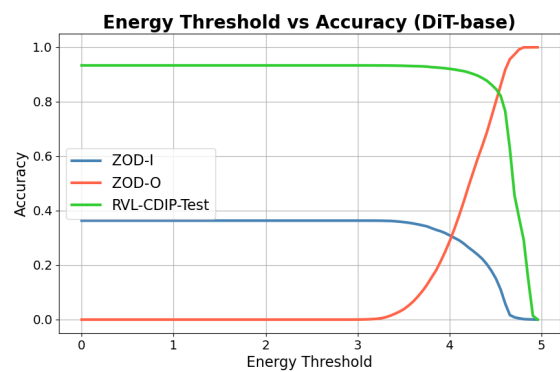


Figure 18. Accuracy scores and density distribution on the basis of energy for DiT-base

3. Examples of ZOD -I

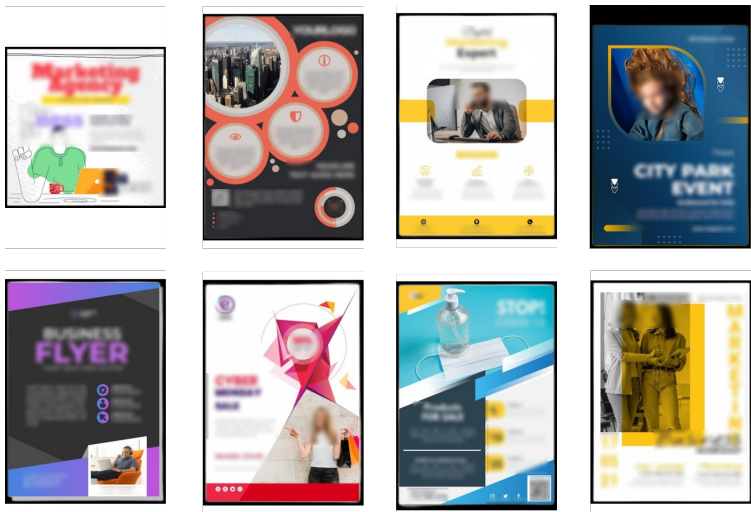


Figure 19. Samples from advertisement class



Figure 20. Samples from file folder class



Figure 21. Samples from presentation class

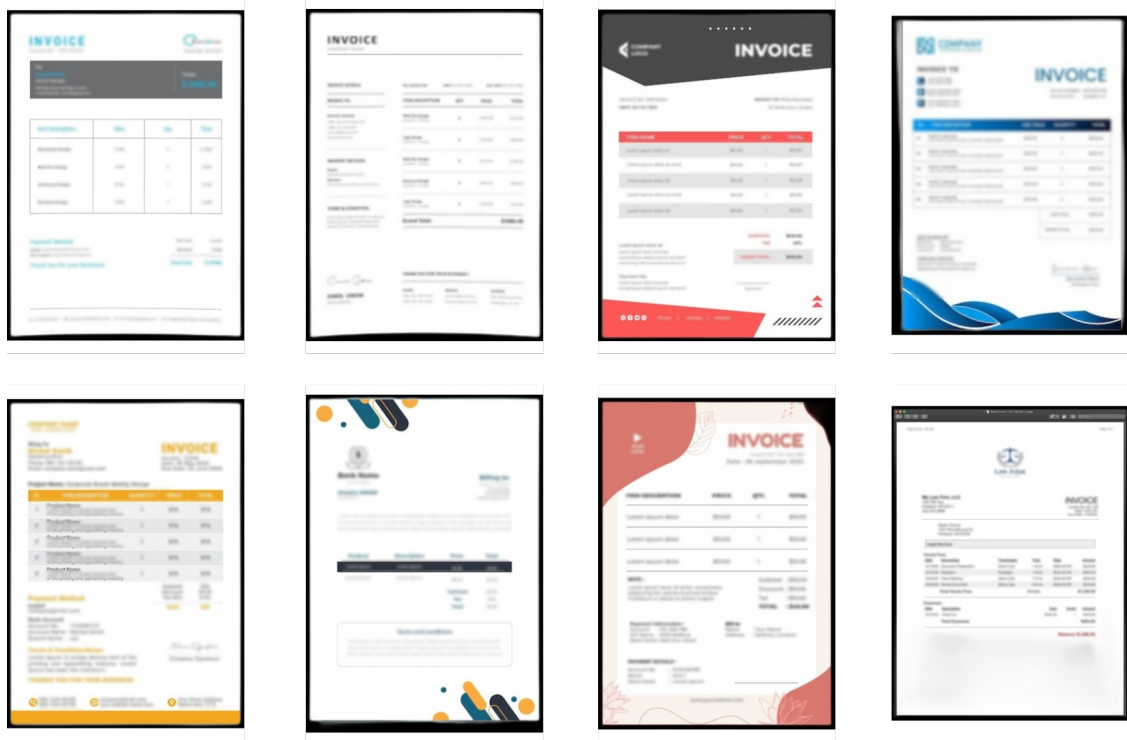


Figure 22. Samples from invoice class

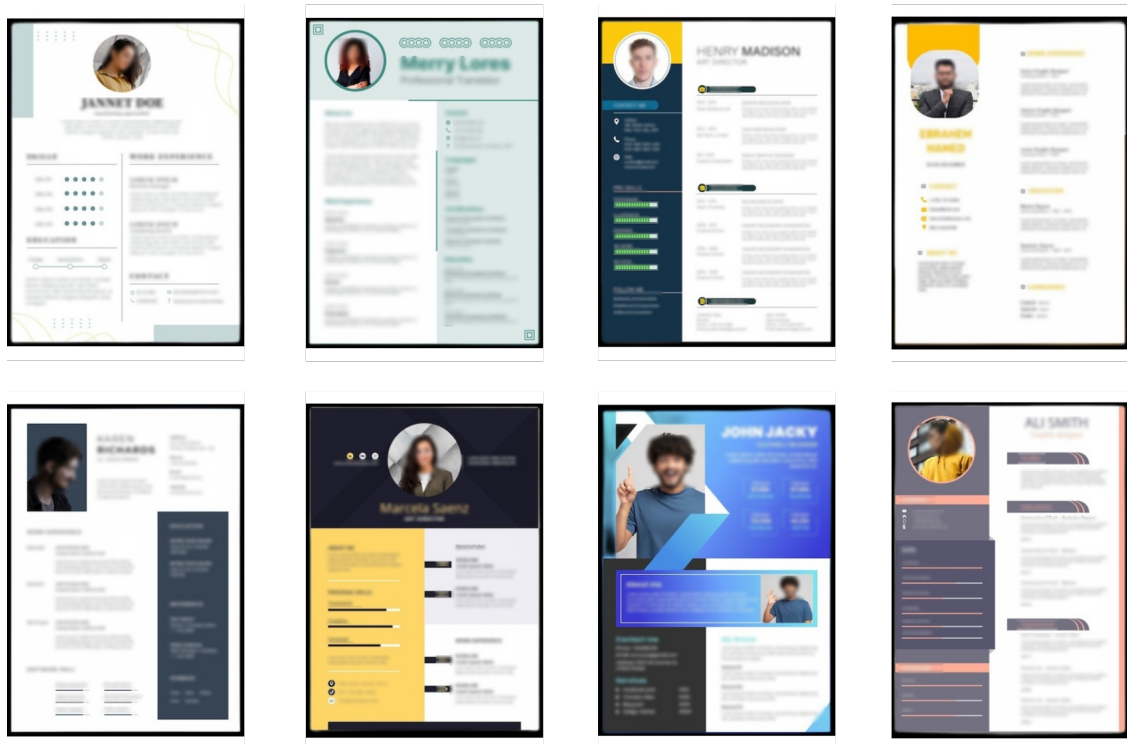


Figure 23. Samples from resume class

4. Examples of ZOD -O

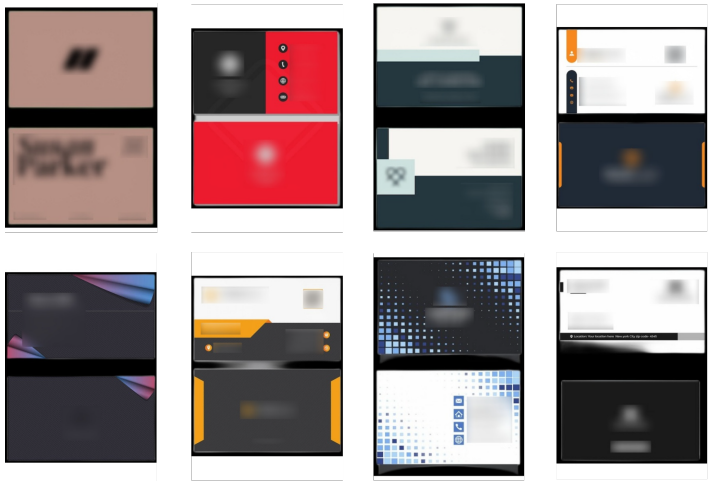


Figure 24. Samples from business cards class

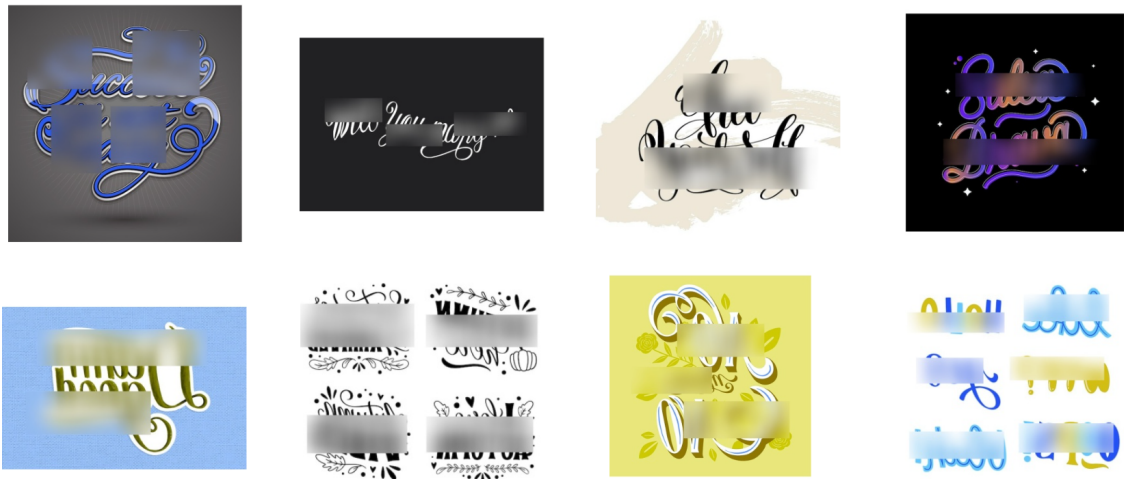


Figure 25. Samples from calligraphy class

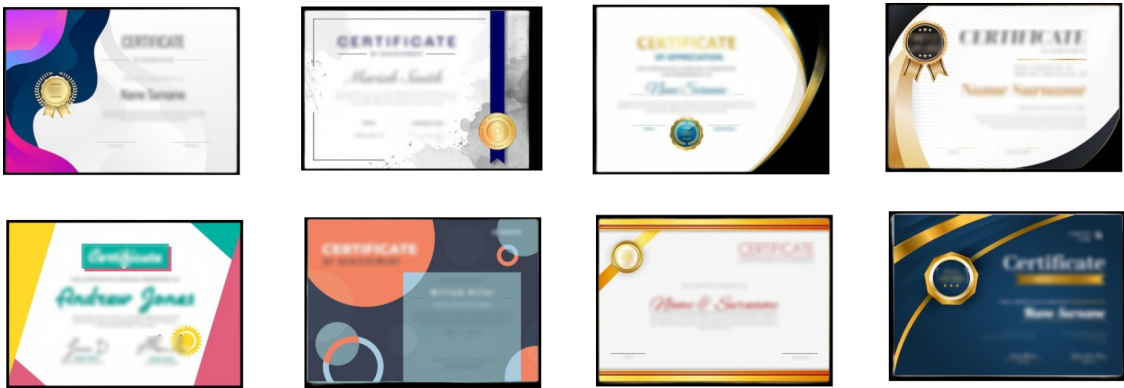


Figure 26. Samples from certificate class

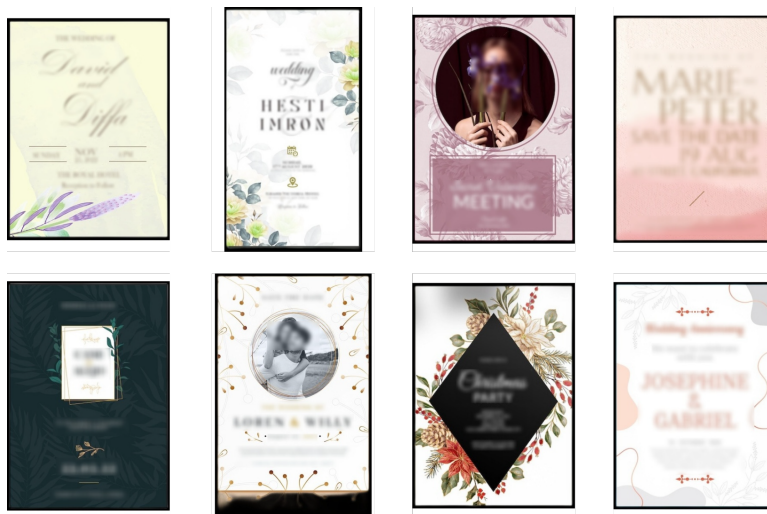


Figure 27. Samples from invitation cards class

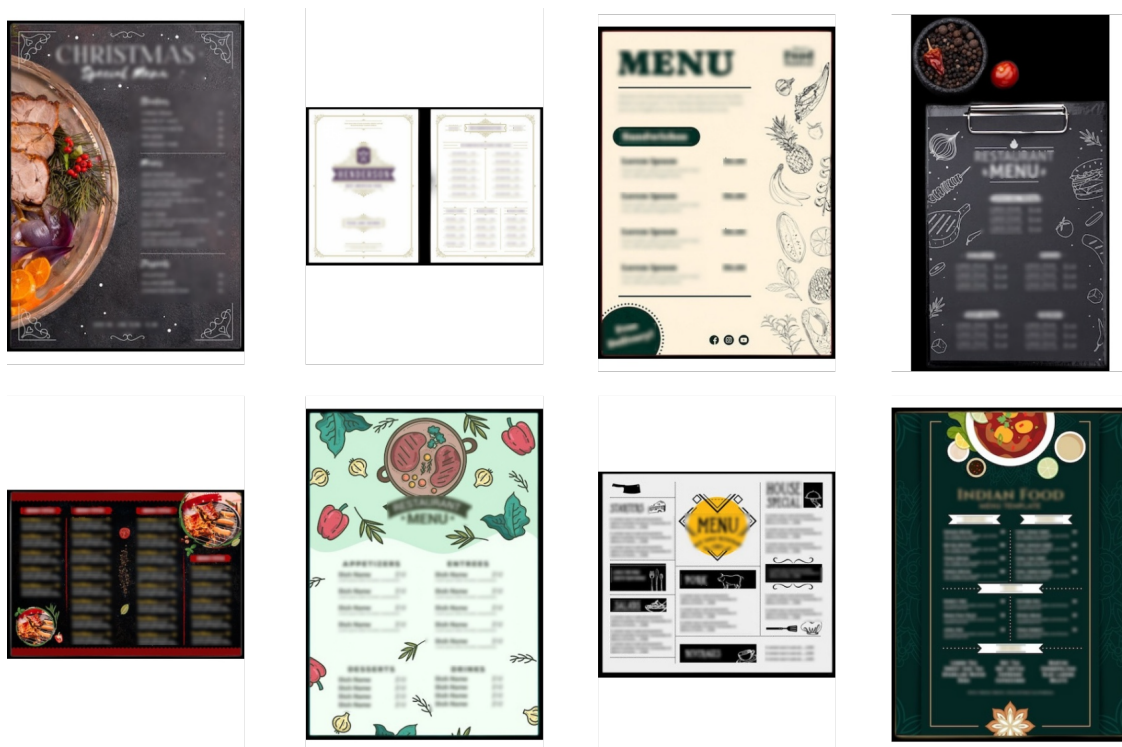


Figure 28. Samples from menu class

5. Examples of ZOD -I with augmentation

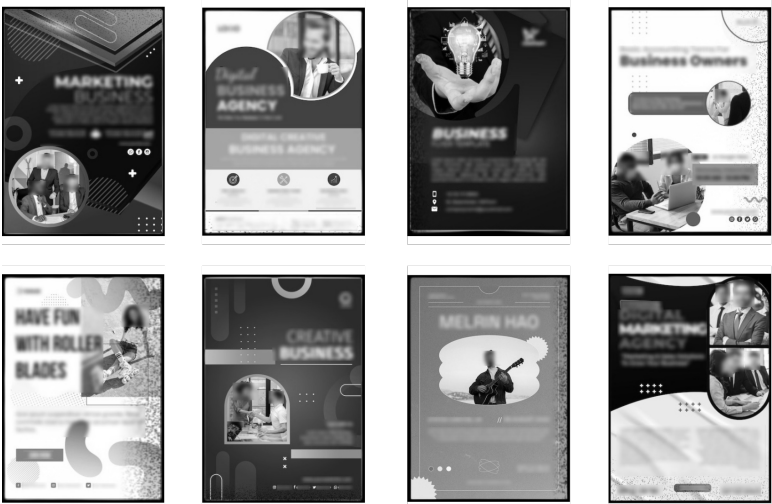


Figure 29. Samples from advertisement class



Figure 30. Samples from invoice class



Figure 31. Samples from presentation class

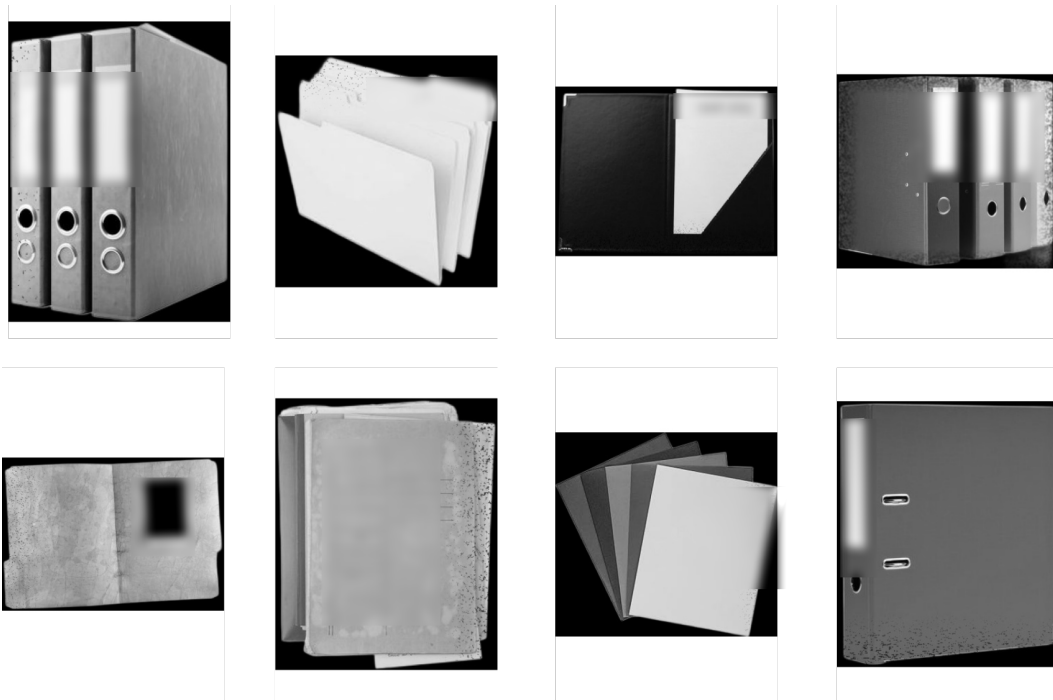


Figure 32. Samples from file folder class



Figure 33. Samples from resume class

6. Examples of ZOD -O with augmentation



Figure 34. Samples from certificate class



Figure 35. Samples from calligraphy class

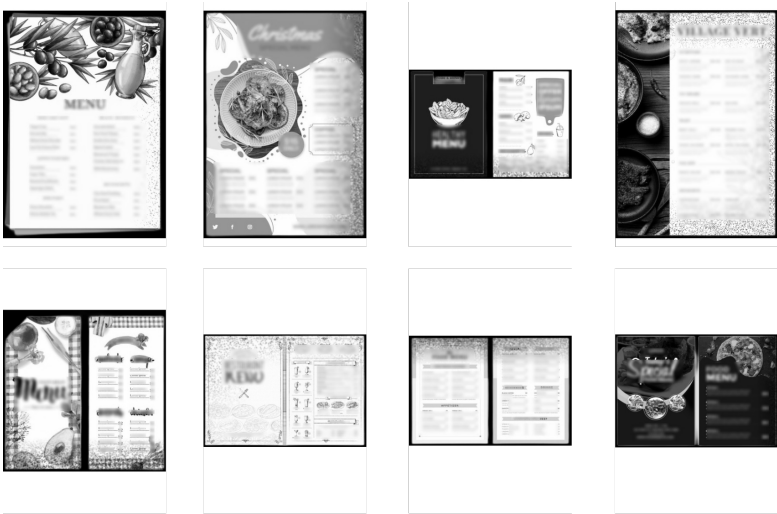


Figure 36. Samples from menu class

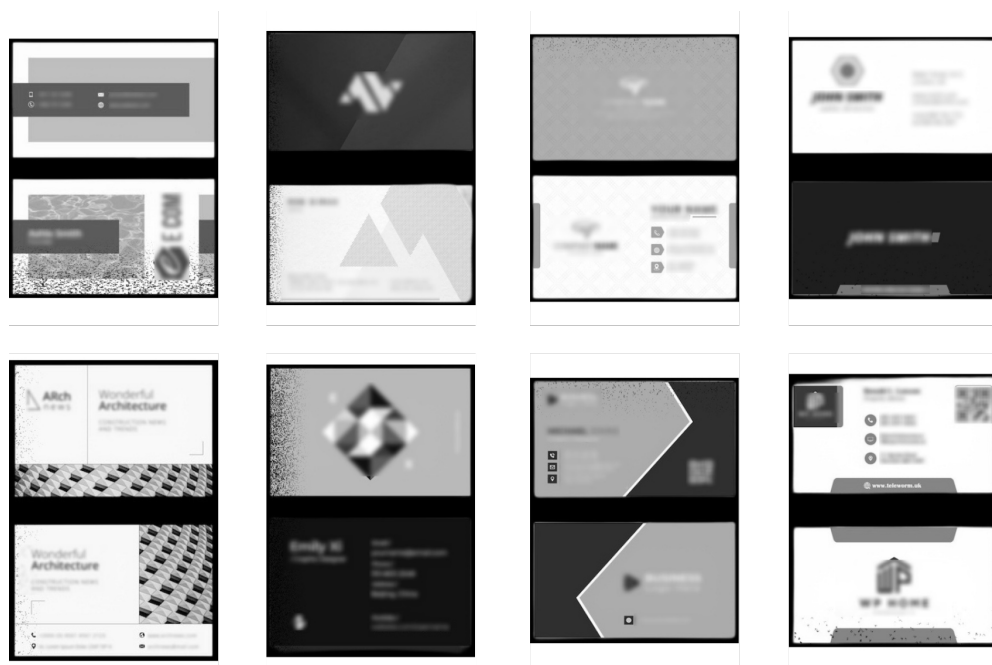


Figure 37. Samples from business cards class



Figure 38. Samples from invitation cards class