TransFusion – A Transparency-Based Diffusion Model for Anomaly Detection

Supplementary material

In this supplementary material, we provide some details and supporting information that extend beyond the scope of the main manuscript and complement it well. Specifically, we show and discuss a number of failure cases, report detailed per-class localisation results, and examine the sensitivity of the proposed method to two hyperparameters. Then, we show the results obtained by fine-tuning hyperparameters to the individual classes, a practice which is used in many recent anomaly detection methods. Finally, we present a comprehensive collection of additional qualitative results, alongside a qualitative comparison with related work, to provide even more insights into the performance of the proposed method.

1 Failure cases

A few failure cases of TransFusion can be seen in Figure 1, where anomalies are not properly localized, or image regions are poorly reconstructed. TransFusion fails to segment tiny anomalous details (Column 1), and outputs masks that do not fit the ground truth in cases (Columns 2 to 5) where it is ambiguous what to annotate as the ground truth. For instance, in Column 2, TransFusion recognizes where the object broke, but the annotators annotate the whole object as anomalous. A similar thing can be noted in Column 3, where the annotators only annotated the hole while the leather around it is curved due to it, which could also be annotated as an anomaly. It also restores the normality of image regions that are relatively out of distribution but are not annotated (Columns 6 and 7). Some of these failure cases impact the anomaly localization score on VisA, where the anomaly masks are small and precise. MVTec AD contains larger anomalies. Therefore, the effect on the anomaly localization score is not as severe. However, the anomaly detection score is impacted.

2 Per-class localization results

Per-class localization results are provided in Table 1 and in Table 2. The lowest scores are achieved for the Fryum and Cashew categories on VisA and for the Transistor and Cable categories on MVTecAD. We hypothesize that this is partly caused by the ambiguous anomalous regions that are difficult to annotate and common in these categories. A few of these ambiguous ground truths can be seen in Figure 1, more specifically in Rows 2, 3, 6 and 7. For instance, in Row 6, TransFusion also reconstructs a part of the shadow that is missing in the original image due to a crack in the cashew. Another example can be seen in

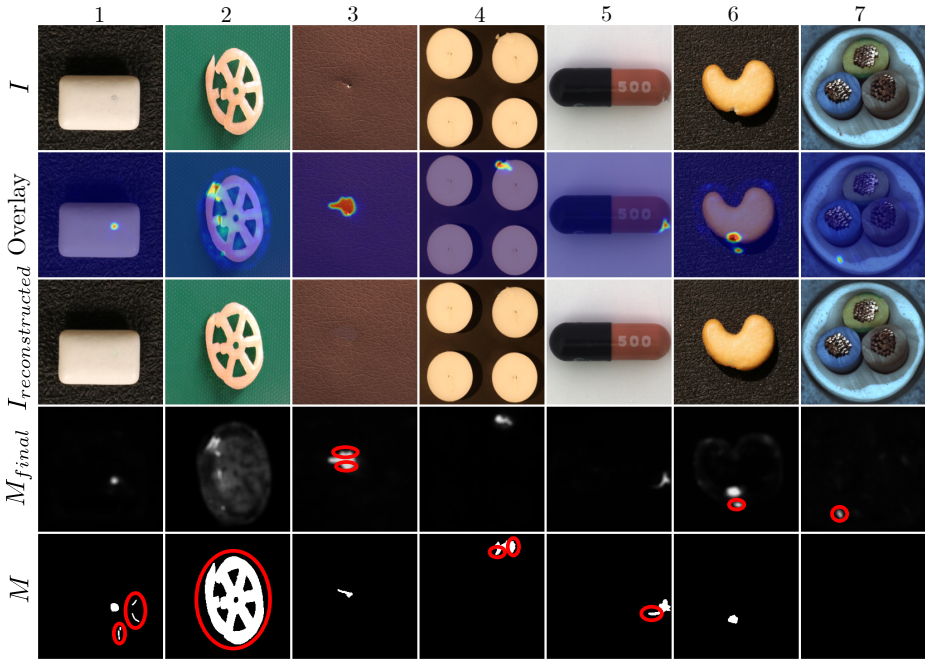

Fig. 1: Failure case results. The anomalous images are shown in the first row, the overlay in the second row, the reconstructions in the third row, the predicted mask, and the real mask in the fourth and fifth rows, respectively. The biggest discrepancies between the predicted and ground truth masks are marked with red circles.

Row 7, where TransFusion fixes a poke in the plastic around the cable. If there are multiple of them in the image this is considered an anomaly in the test set, so the annotation of this image is ambiguous.

Category	Candle	Capsules	Cashew	Chewing gum	Fryum	Macaroni1	Macaroni2	PCB1	PCB2	PCB3	PCB4	Pipe fryum	*Average*
TransFusion	88.6	97.3	82.8	83.2	77.8	94.0	95.6	92.4	85.1	92.0	89.4	87.9	88.8

Table 1: Detailed results for Transfusion for anomaly localization on VisA. All results are reported in AUPRO.

3 Additional ablation study results

Weight size. In the final mask calculation (Eq. (10)) the weight λ defines the impact of M_{disc} and M_{recon} on the final mask. TransFusion's performance under various λ values is shown in Figure 2. The results are robust for larger λ values,

Category	Carpet	Grid	Leather	Tile	Wood	Bottle	Cable	Capsule	Hazelnut	Metal nut	Pill	Screw	Toothbrush	Transistor	Zipper	*Average*
TransFusion	95.9	98.0	96.2	95.0	94.8	97.3	85.5	92.1	97.7	94.1	96.2	97.0	94.1	83.9	97.2	94.3

Table 2: Detailed results for Transfusion for anomaly localization on MVTec AD. All results are reported in AUPRO.

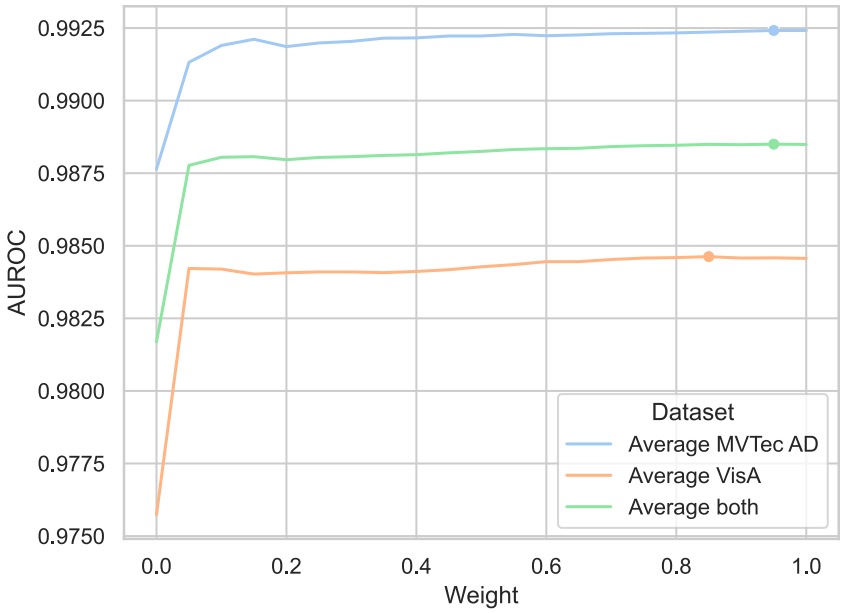

Fig. 2: Average AUROC for different weights λ in the final mask calculation. The maximum point of each line is represented with a dot.

where M_{disc} has a higher impact on the final mask. However, the best results are achieved with λ values at which M_{recon} still impacts the final mask.

Kernel size. To determine the final mask calculation as described in Eq. (10), we incorporated a mean filter f_n of size n into the formulation. Here we explore TransFusion's behaviour under various values of n. The results can be seen in Figure 3. Note that higher values of n quickly deteriorate the performance on the VisA dataset due to the scale of anomalies present in the dataset. On the MVTec AD dataset, high kernel sizes have little to no effect on the anomaly detection performance.

4 Per-class tuned results for anomaly detection

Some recent works [13, 21, 22] report performance where hyperparameter tuning was done for each class individually. We maintain a single set of hyperparameters

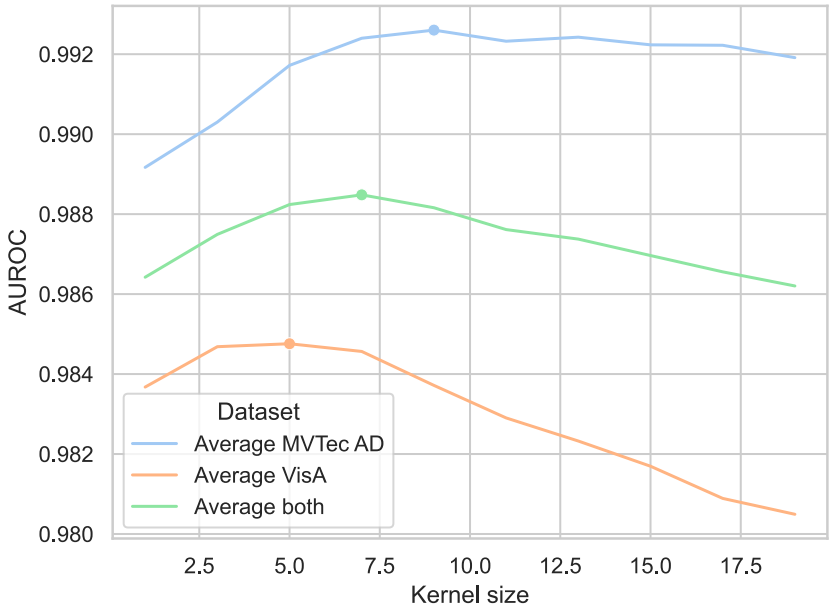

Fig. 3: Average AUROC for different kernel sizes n in the final mask calculation. The maximum point of each line is represented with a dot.

Category	Candle	Capsules	Cashew	Chewing gum	Fryum	Macaroni1	Macaroni2	PCB1	PCB2	PCB3	PCB4	Pipe fryum	*Average*
Detection	98.3	99.7	96.8	99.9	98.7	99.4	96.8	99.1	99.9	99.5	99.6	99.8	99.0

Table 3: Best possible results for TransFusion when we choose the optimal number of epochs for each class on VisA. Anomaly detection results are reported in AUROC.

for all experiments in the paper. For instance, the total number of epochs was set in stone, and the result was calculated using the model from the final epoch. For the sake of completeness, we report results where the total number of epochs was optimized for each class. These results enable future works to be compared with per-class tuned models. The results are shown in Table 3 and Table 4. Results on VisA [45] exceed the current highest score by 0.9%, and results on MVTec AD [6] improve even further.

5 Additional qualitative results

In this section, we provide more qualitative results. Figure 4 and Figure 5 show some result samples from each category on both datasets. As we can observe, TransFusion outputs very precise masks that closely match the ground truth annotation in the vast majority of cases.

Category	Carpet	Grid	Leather	Tile	Wood	Bottle	Cable	Capsule	Hazelnut	Metal nut	Pill	Screw	Toothbrush	Transistor	Zipper	*Average*
Detection	99.8	100	100	100	99.9	100	98.4	98.8	100	100	99.5	97.2	100	98.8	100	99.5

Table 4: Best possible results for TransFusion when we choose the optimal number of epochs for each class on MVTec AD. Anomaly detection results are reported in AUROC.

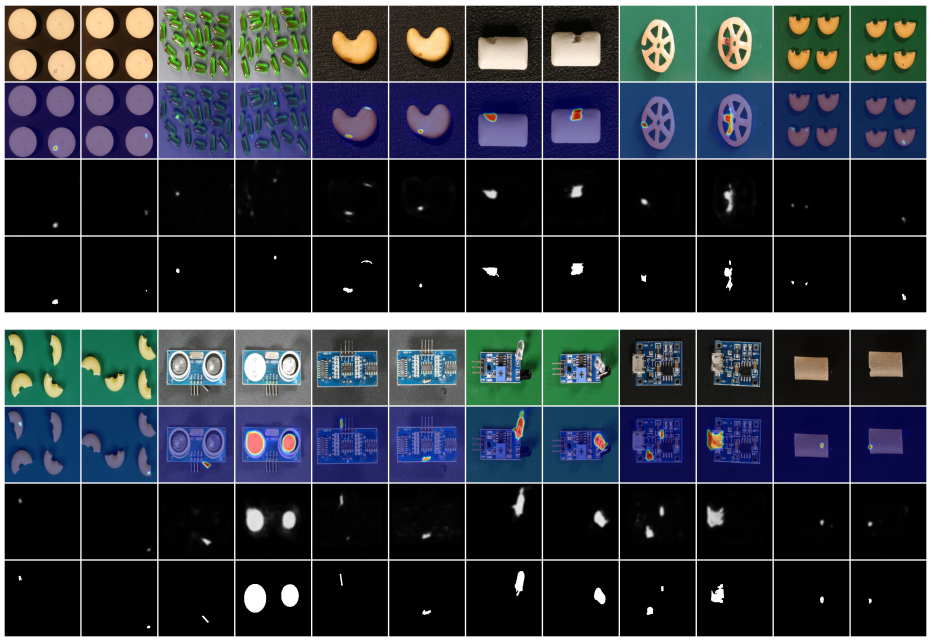

Fig. 4: Qualitative examples on VisA dataset. The original image, the anomaly map overlay, the anomaly map and the ground truth map are shown.

6 Additional qualitative comparisons to other methods

This section provides more qualitative mask comparisons to other state-of-the-art methods. We compared TransFusion with DRAEM [40], RD4AD [11], Patchcore [25] and DiffAD [43]. The results can be seen in Figure 6.

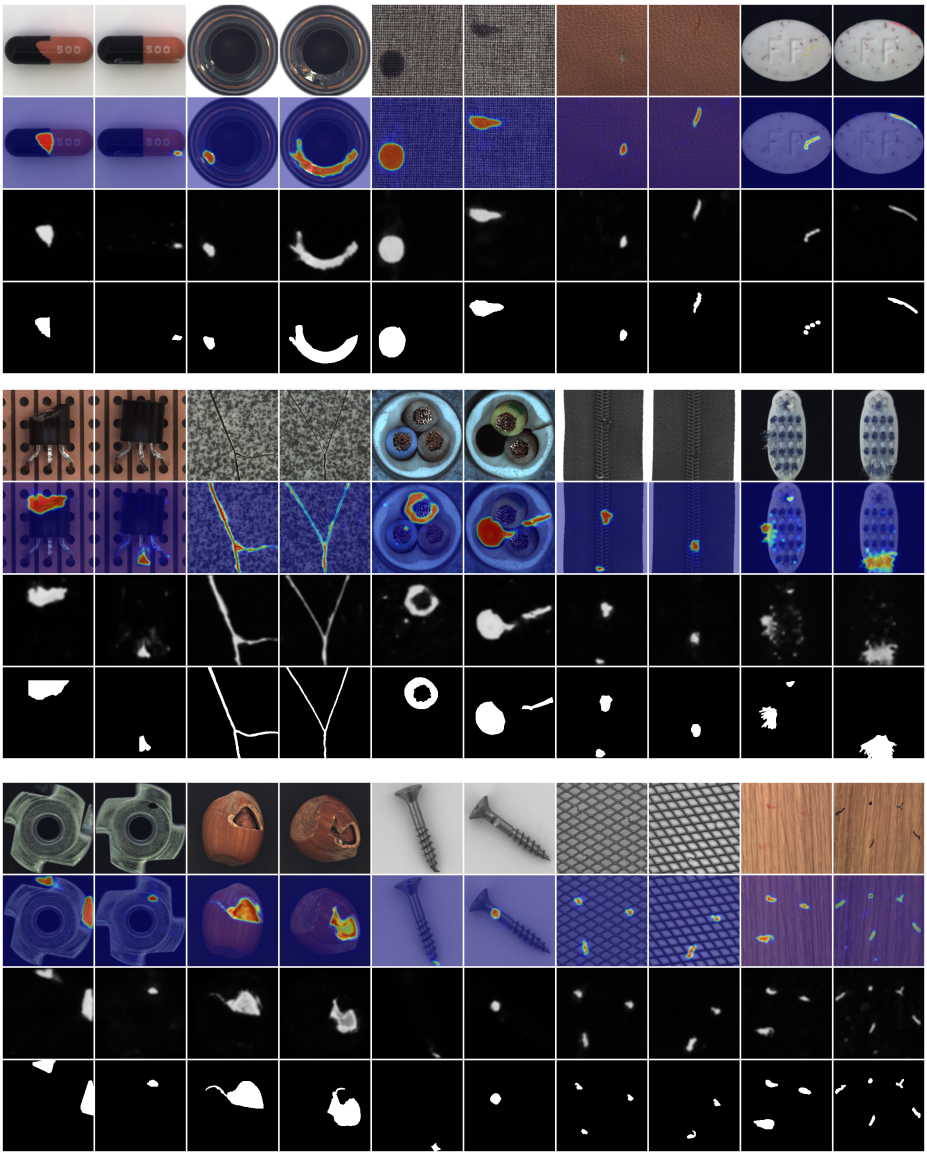

Fig. 5: Qualitative examples on MVTec AD dataset. The original image, the anomaly map overlay, the anomaly map and the ground truth map are shown.

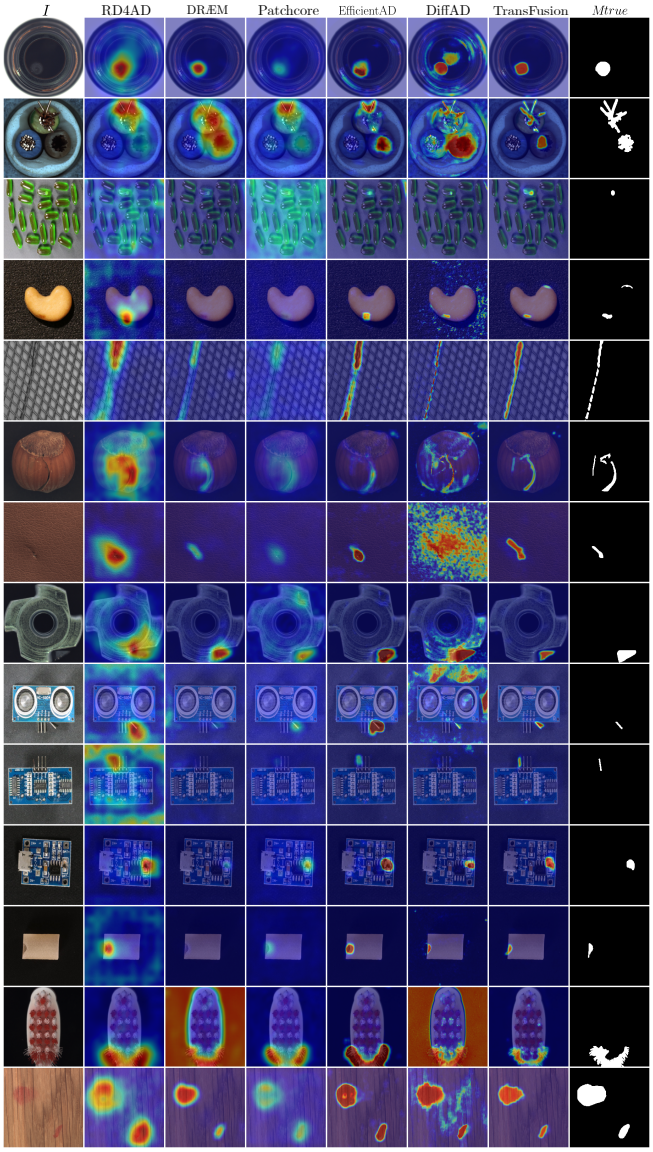

Fig. 6: Qualitative comparison of the masks produced by TransFusion and five other state-of-the-art methods. The anomalous images are shown in the first column. The middle six columns show the anomaly mask generated by RD4AD [11], DRÆM [40], Patchcore [25], EfficientAD [5], DiffAD [43] and TransFusion respectively. The last column shows the ground truth anomaly mask.