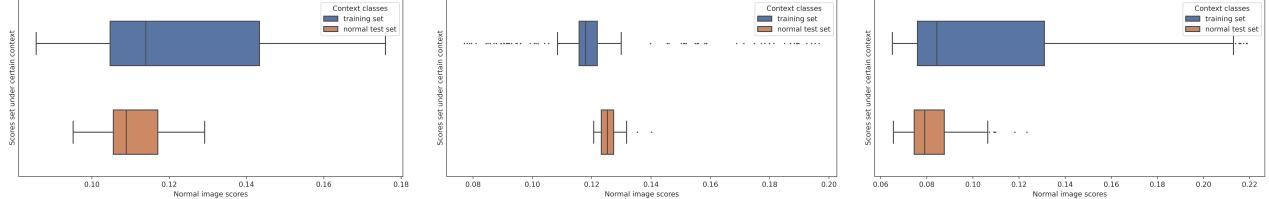
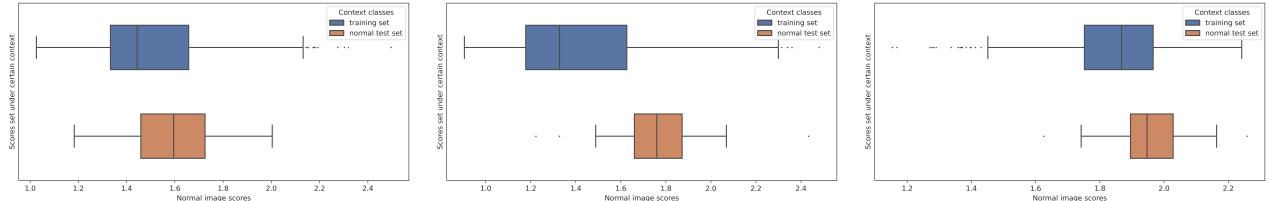


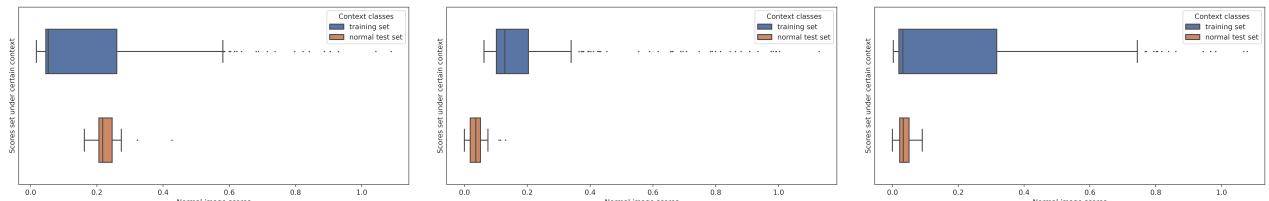
## Hierarchical Martingale-based Testing Framework for Visual Anomaly Detection



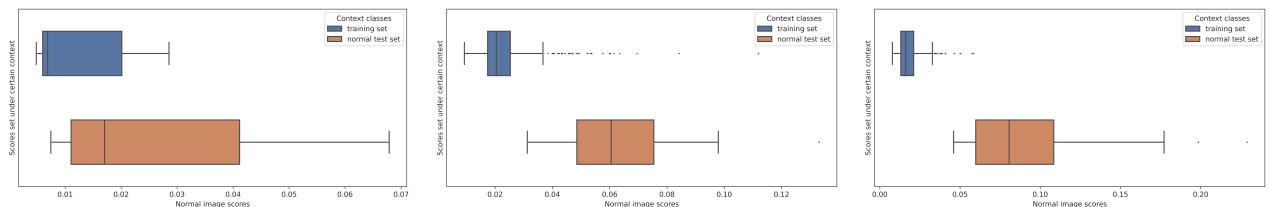
(a) CFA-BTAD-01; CFA-MVTec-Leather; CFA-VisA-Candle (from left to right).



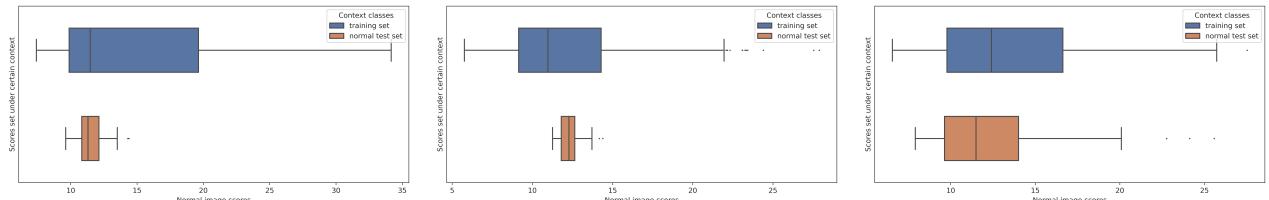
(b) CFLOW-BTAD-02; CFLOW-MVTec-Wood; CFLOW-VisA-Fryum (from left to right).



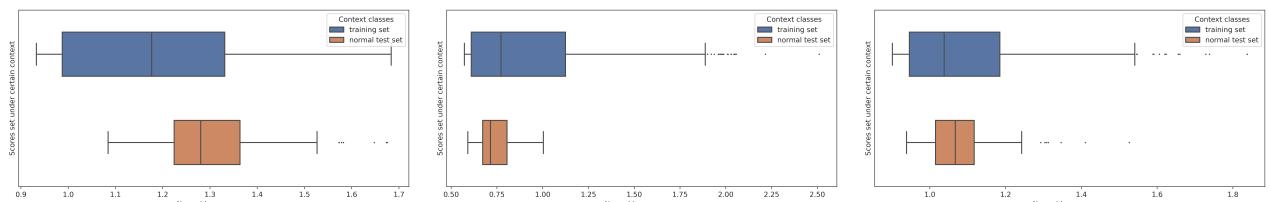
(c) DMAD-BTAD-02; DMAD-MVTec-Carpet; DMAD-VisA-Fryum (from left to right).



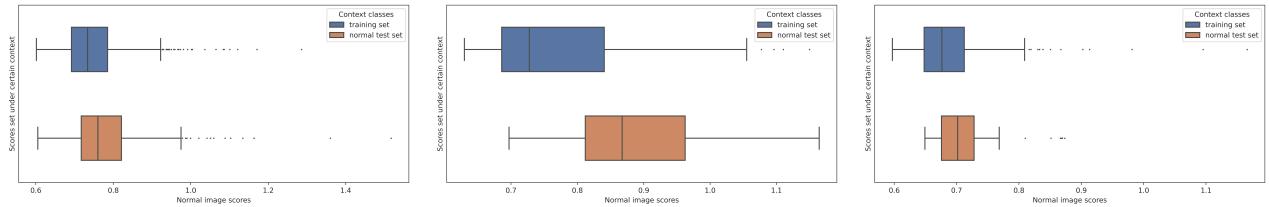
(d) DRAEM-BTAD-01; DRAEM-MVTec-Metal\_nut; DRAEM-VisA-Cashew (from left to right).



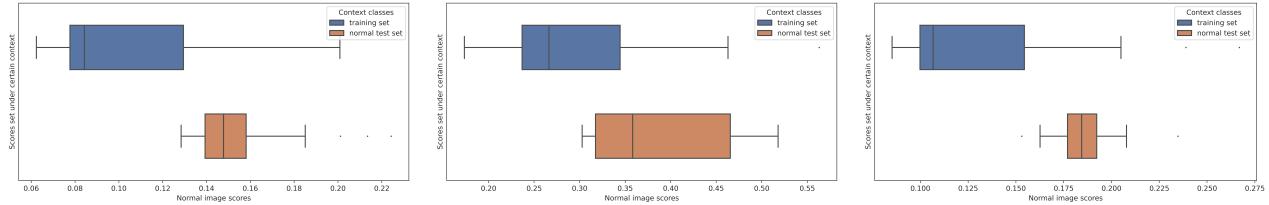
(e) PaDiM-BTAD-02; PaDiM-MVTec-Leather; PaDiM-VisA-Candle (from left to right).



(f) PatchCore-BTAD-03; PatchCore-MVTec-Capsule; PatchCore-VisA-Pcb2 (from left to right).

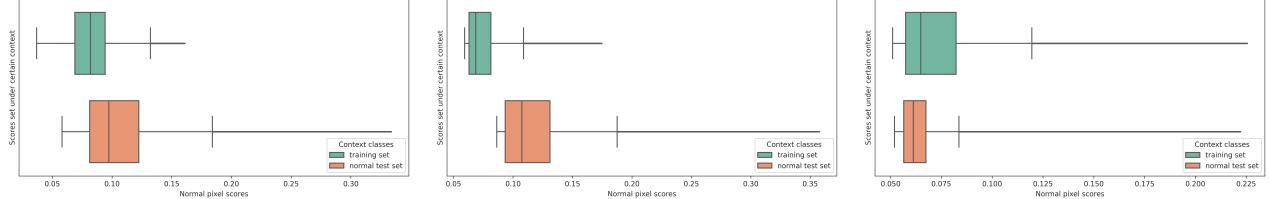


(g) ReContrast-BTAD-03; ReContrast-MVTec-Transistor; ReContrast-VisA-Pipe\_fryum (from left to right).

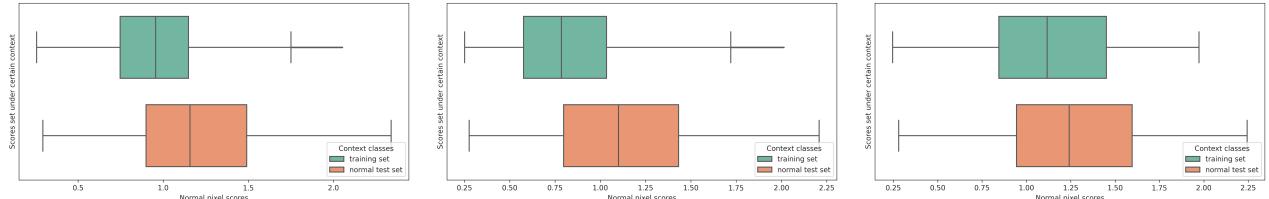


(h) RRD-BTAD-02; RRD-MVTec-Toothbrush; RRD-VisA-Chewinggum (from left to right).

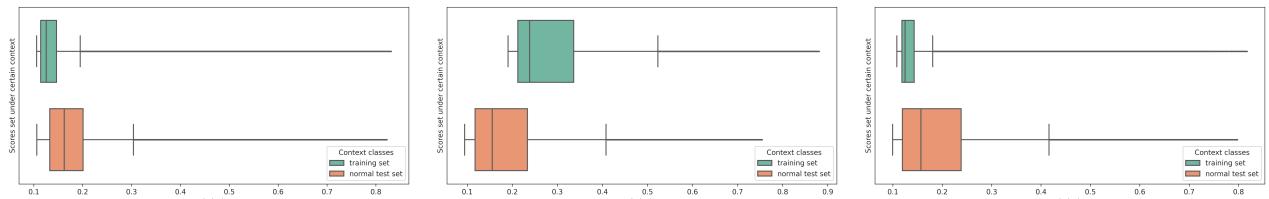
Figure 10: Image-level distribution shifts in normal scores. The term *CFA-MVTec-Leather* indicates that scores are generated by CFA on Leather category of MVTec dataset. For instance, the upper plot in (a) denotes the distribution in normal image scores of training set, and the bottom plot in (a) denotes the distribution in normal image scores of test set.



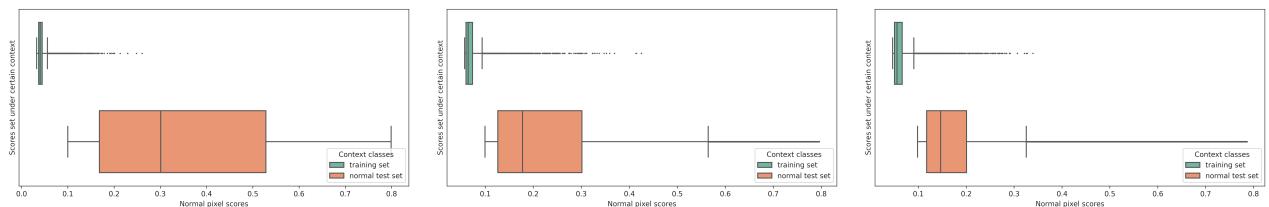
(a) CFA-BTAD-01; CFA-MVTec-Leather; CFA-VisA-Candle (from left to right).



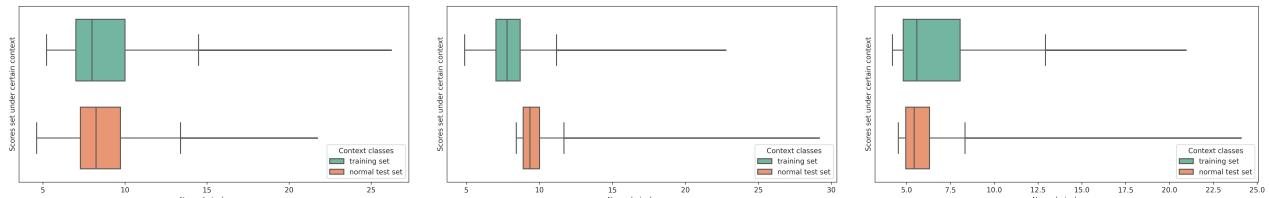
(b) CFLOW-BTAD-02; CFLOW-MVTec-Wood; CFLOW-VisA-Fryum (from left to right).



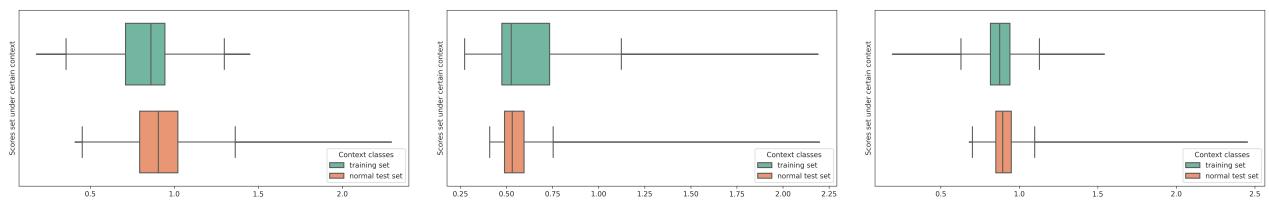
(c) DMAD-BTAD-02; DMAD-MVTec-Carpet; DMAD-VisA-Fryum (from left to right).



(d) DRAEM-BTAD-01; DRAEM-MVTec-Metal\_nut; DRAEM-VisA-Cashew (from left to right).



(e) PaDiM-BTAD-02; PaDiM-MVTec-Leather; PaDiM-VisA-Candle (from left to right).



(f) PatchCore-BTAD-03; PatchCore-MVTec-Capsule; PatchCore-VisA-Pcb2 (from left to right).

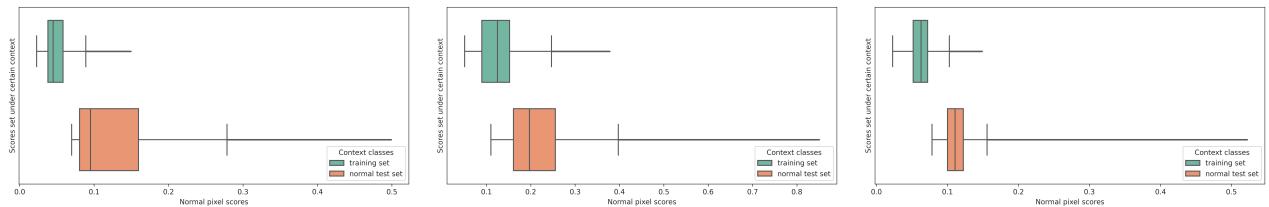
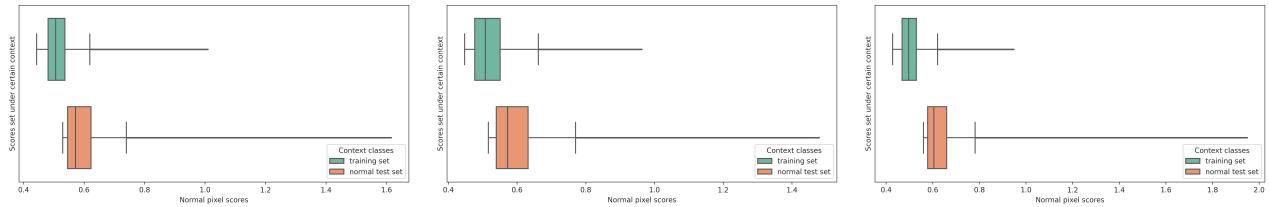


Figure 11: Pixel-level distribution shifts in normal scores. The term *CFA-MVTec-Leather* indicates that scores are generated by CFA on Leather category of MVTec dataset. For instance, the upper plot in (a) denotes the distribution in normal image scores of training set, and the bottom plot in (a) denotes the distribution in normal image scores of test set.