Computer Engineering Department

A Survey on the Accuracy and Performance of Video Anomaly Detection Models

Project Advisor: Haonan Wang

Gangasani, Lokesh Reddy (MS Software Engineering)
Juvvadi, Mounish (MS Software Engineering)
Kondreddy, Lohith Kumar Reddy (MS Software Engineering)
Pasala, Santosh Sai Gowtham (MS Software Engineering)

Introduction

Anomaly detection in computers involves the identification of unusual patterns or deviations in data that may signify potential threats. In this study, we delve into the realm of video anomaly detection, a crucial area in video surveillance systems driven by security concerns and advancements in machine learning. Our objective is to survey four leading models in video anomaly detection, each employing distinct methodologies to tackle the inherent challenges. These models include Attention-based residual autoencoder for video anomaly detection (ASTNET), Attribute-based Representations for Accurate and Interpretable Video Anomaly Detection (AIVAD), Fast Anomaly Detection via Spatio-temporal Patch Transformation (FASTANO), and Diversity-Measurable Anomaly Detection (DMAD), which leverage deep representations, attention mechanisms and spatial-temporal transformations to enhance accuracy and efficiency.

To understand the differences among the four forementioned approaches, we design and conduct a list of experiments. From the experimental results, we analyze and compare these models by meticulously examining their methodologies, assessing their accuracy using key metrics such as AUC, PSNR, CPU time and measuring their performance on training and inference across different dataset sizes (100% to 12%) and frame sizes(128x128 and 256x256). Through our comparative analysis, we aim to unravel the strengths and limitations of each model, shedding light on their performance.

Our contributions are as follows:

- We conduct a survey of four models and provide detailed information about each.
- We perform experiments under diverse conditions.

Background

Deep learning for anomaly detection relies on neural networks, which are composed of interconnected nodes organized into layers and utilize weighted connections. Autoencoders, a type of neural network, are particularly useful for unsupervised learning tasks such as anomaly detection, as they can compress data and then reconstruct it to quantify anomalies. Attention mechanisms enhance model performance by allowing it to focus on specific parts of input data, improving interpretability and efficiency, even with limited computing resources.

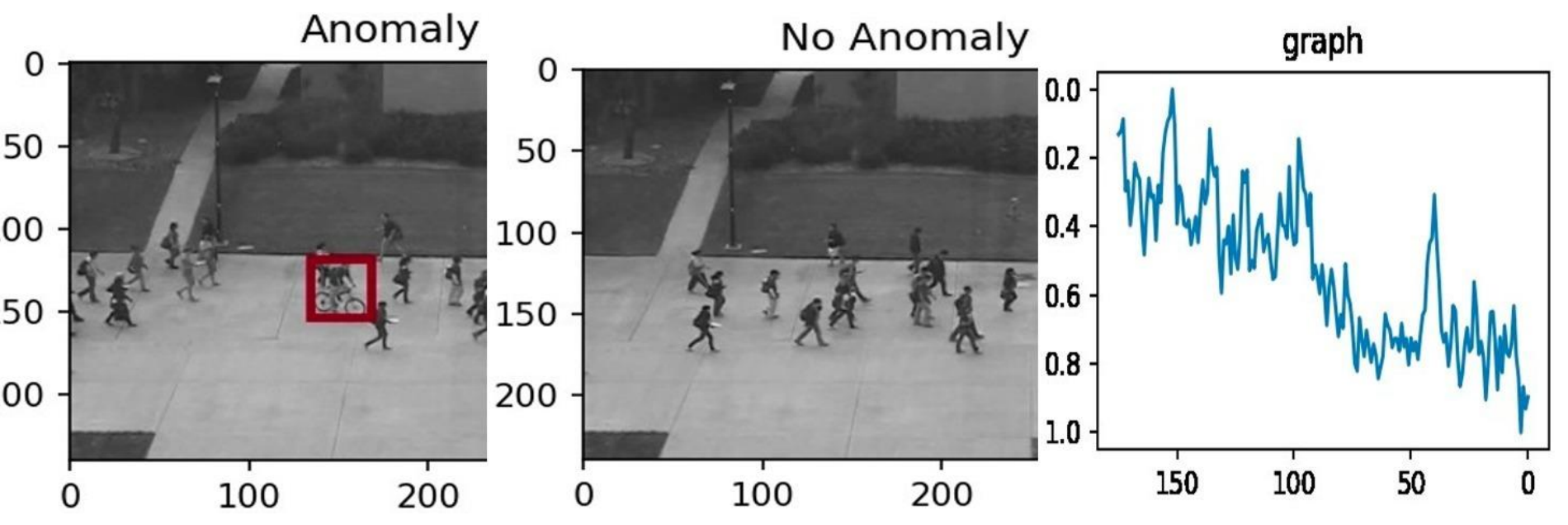

Fig 1: Anomaly Detection Visualization

DMAD (Diversity Measurable Anomaly Detection): The DMAD framework incorporates an Information Compression Module (ICM) and a Pyramid Deformation Module (PDM) to enhance anomaly detection performance. The ICM creates prototypical embeddings of normal patterns, while the PDM models diverse normal patterns and measures anomaly severity. This integration improves anomaly discrimination and reconstruction diversity.

AI-VAD (Attribute-based Representations for Accurate and Interpretable Video Anomaly Detection): The methodology for AI-VAD involves three stages: pre-processing, feature extraction, and density estimation. During the pre-processing stage, optical flow maps and bounding boxes are extracted to represent objects and motions. AI-VAD combines these explicit attribute-based representations with implicit deep representations. By leveraging semantic attributes like velocity and pose, along with deep features, the model offers high interpretability and accuracy in detecting anomalies, making it suitable for critical environments.

FASTANO (Fast Anomaly Detection via Spatio-temporal Patch Transformation): FASTANO utilizes spatial rotation transformation (SRT) and temporal mixing transformation (TMT) to enhance the learning of normal features for fast anomaly detection in surveillance videos. This approach, which involves patch anomaly generation and an autoencoder architecture, focuses on moving objects and does not rely on pre-trained networks. The model is trained to minimize prediction loss and generate frames resembling the ground truth, achieving competitive results on three datasets.

ASTNET (Attention-based residual autoencoder for video anomaly detection): ASTNET leverages a residual autoencoder architecture with a deep convolutional neural network-based encoder and a multi-stage channel attention-based decoder for video anomaly detection. By effectively combining spatial and temporal features using channel attention modules and a temporal shift method, it outperforms existing methods on standard benchmark datasets. The architecture features an encoder-decoder structure, with the encoder capturing appearance and motion details through convolutional neural network, such as WiderResnet. The temporal branch employs a novel shifting technique to model temporal features across frames, while the spatial branch aggregates features from multiple frames.

Evaluation Methodology and Results

Experimental Setup

We conduct experiments on system with AMD Ryzen 7 7745HX with Radeon Graphics, 3601 MHz, 8 Core(s), 16 Logical Processor(s), CPU - 16GB RAM 5.15GHz clock cycle, Graphics - NVIDIA® GeForce RTX™ 4060 8GB GDDR6 8.00GHz, Memory number of channels - 2*32-bit, Peak Bandwidth - 2800 MHz, Bus Type Pci-express 5.0 (32.0) gt/s 99.79 MHz. We evaluate performance across various datasets, data sizes, and frame sizes using metrics such as AUC, PSNR scores, and CPU times.

Metrics

AUC (Area Under Curve): AUC provides a comprehensive measure of a model's ability to distinguish between classes, specifically the presence or absence of anomalous activity. We assess the model's accuracy in detecting anomalies by comparing the labels assigned to generated frames with ground truth annotations.

PSNR (Peak Signa to Noise Ratio): PSNR calculates the quality of reconstructed frames compared to original frames and used to assess the distortion caused by anomalies.

CPU Time: It refers to the amount of time the CPU spends processing data to detect anomalies, indicating the efficiency of the algorithm.

Datasets

PED2: The Ped2 dataset comprises 16 training and 12 testing video samples for developing surveillance anomaly detection algorithms.

Avenue: The Avenue Dataset consists of 16 training and 21 testing video clips, totaling 30,652 frames captured on the CUHK campus avenue, with 15,328 frames allocated to the training set and 15,324 to the testing set.

Shanghai Tech: The Shanghai Tech Campus dataset features 13 scenes with diverse lighting conditions and camera angles, providing over 270,000 training frames for extensive anomaly detection

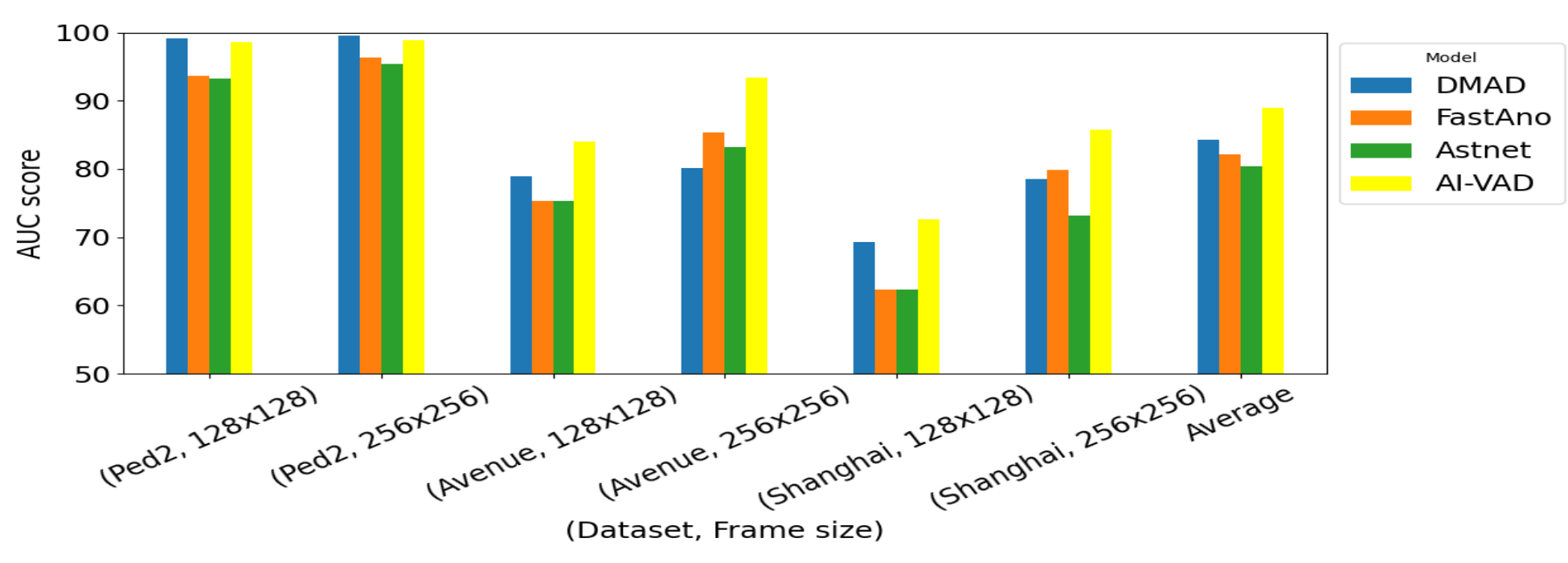

Fig 2: AUC Values of Various Models by Dataset and Frame Size for 100% Dataset

We observe that in Figure 2, DMAD model has the highest AUC score of 99.14 in the PED2 dataset and ASTNET has the lowest. But when the dataset changed to Avenue and Shanghai, AI-VAD model has the highest AUC Score and ASTNET has the lowest. On average, AI-VAD has the highest AUC and ASTNET has the lowest.

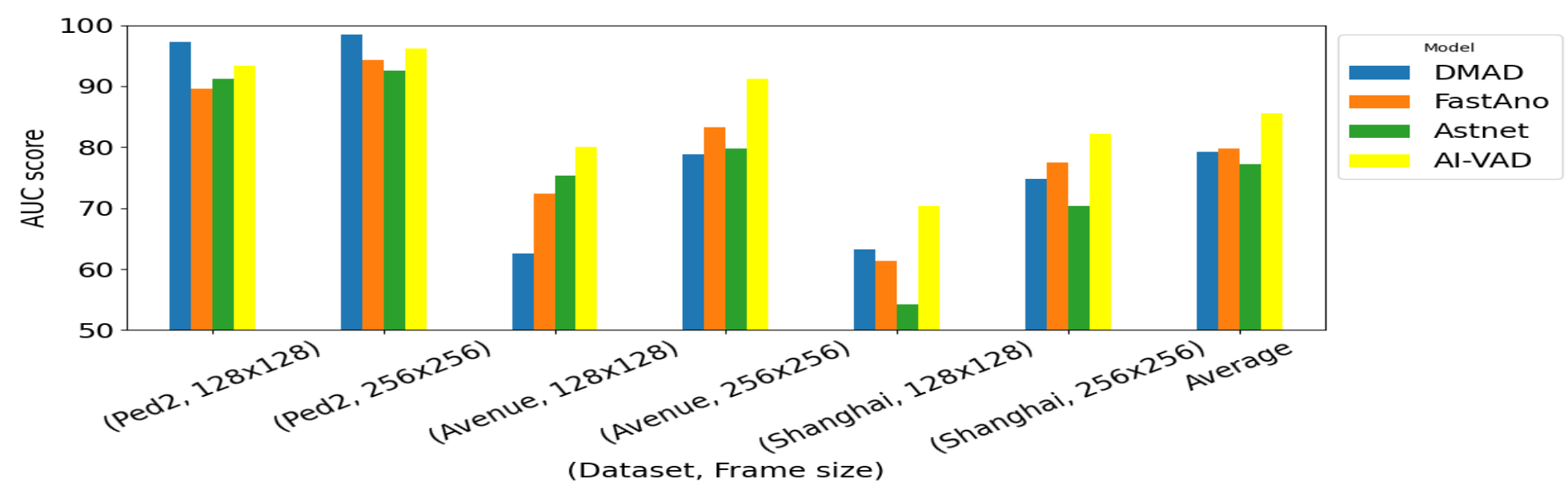

Fig 3: AUC Values of Various Models by Dataset and Frame Size for 25% Dataset

We observe that in Figure 3, when the dataset is reduced to 25%, there is a significant decrease in AUC scores. Despite this, the DMAD model still achieves the highest AUC score on the PED2 dataset, whereas the ASTNET model has the lowest. Also, the ASTNET model is sensitive to the Shanghai tech dataset for the frame size of 128x128, where it has the lowest AUC score. Overall, the AI-VAD model exhibits highest average AUC and ASTNET model has lowest.

Also, when the dataset size is reduced from 100% to 25%, DMAD's average score falls below that of FastAno, while AI-VAD remains superior in both cases. Additionally, across different datasets and frame sizes, AI-VAD consistently achieves the highest average AUC score compared to ASTNET, which consistently performs lower.

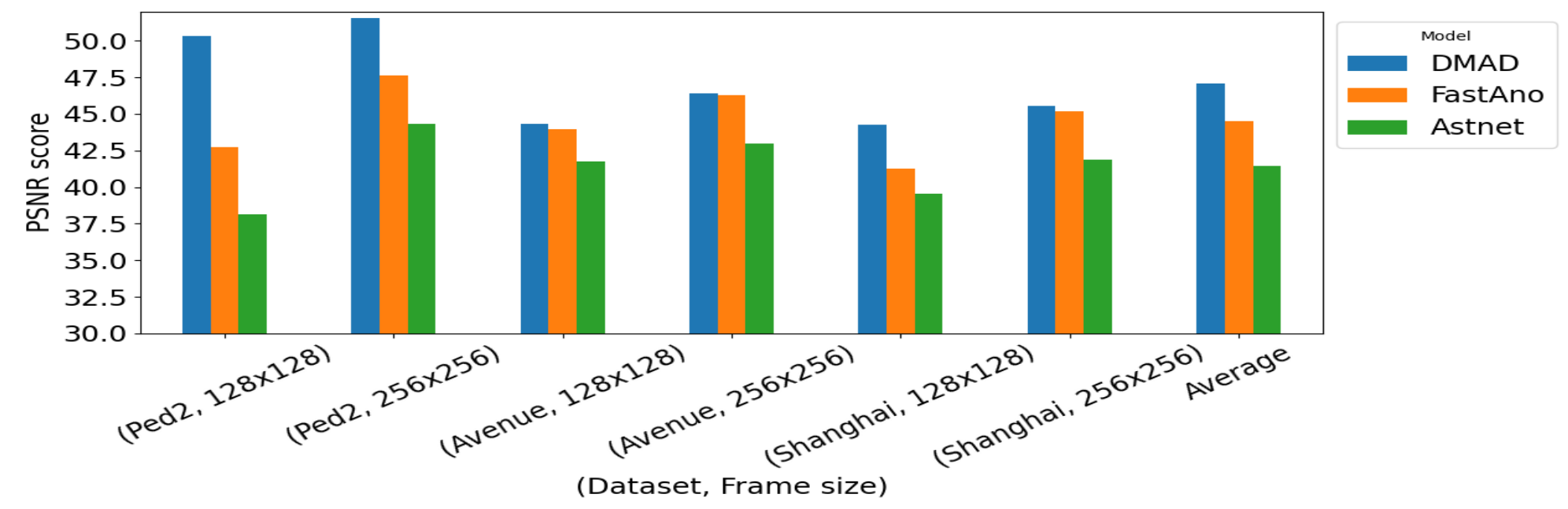

Fig 4: PSNR scores of various models by dataset and frame size for 100% dataset

We observe that in Figure 4, PSNR scores for 100% dataset are represented and DMAD model has the highest PSNR score and ASTNET has the lowest. But when the dataset and frame sizes are changed, DMAD values are almost equal to FASTANO. On average, DMAD has the highest PSNR and ASTNET has the lowest.

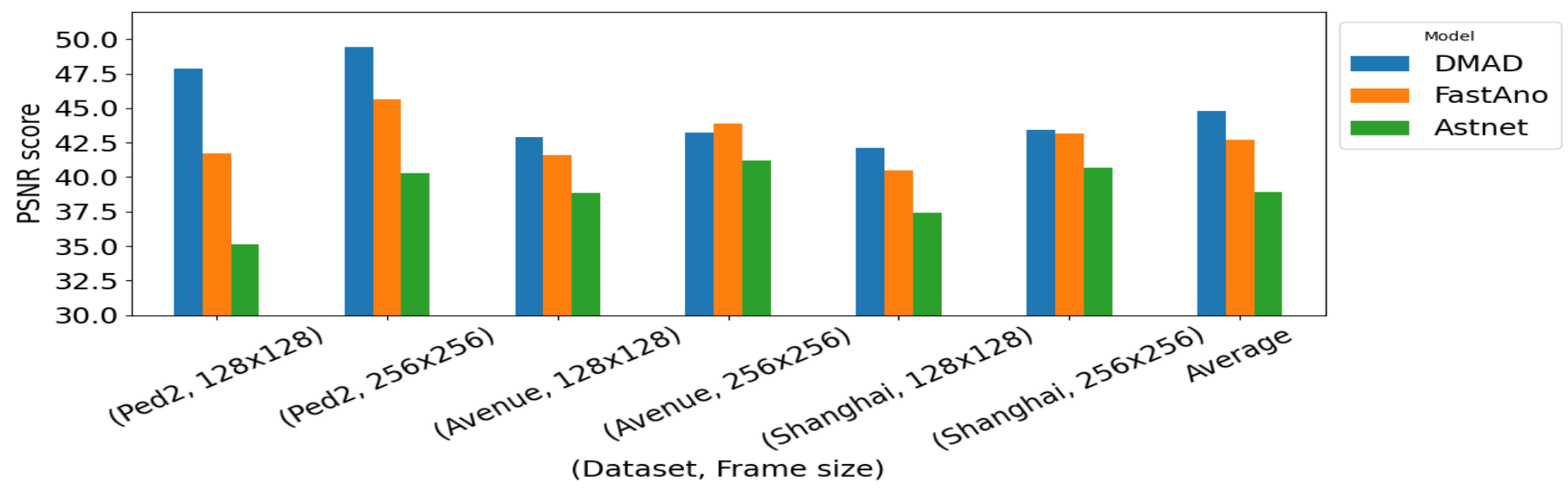

Fig 5: PSNR scores of various models by dataset and frame size for 25% dataset

We observe that in Figure 5, As the dataset is reduced to 25% the PSNR scores are slightly affected. DMAD model has the highest PSNR score and ASTNET has the lowest in PED2 dataset. But when the dataset and frame sizes are changed, FASTANO shown highest PSNR in Avenue dataset with frame size 256x256. On average, AI-VAD has the highest PSNR and ASTNET has the lowest.

When the dataset size is reduced to 25%, DMAD maintains its superiority with the highest PSNR score in the Ped2 dataset, although FastAno outperforms DMAD for Avenue at 256x256 frame size. In AI-VAD, we can't calculate PSNR since no frames are generated. Instead, we evaluate the velocity and pose representations.

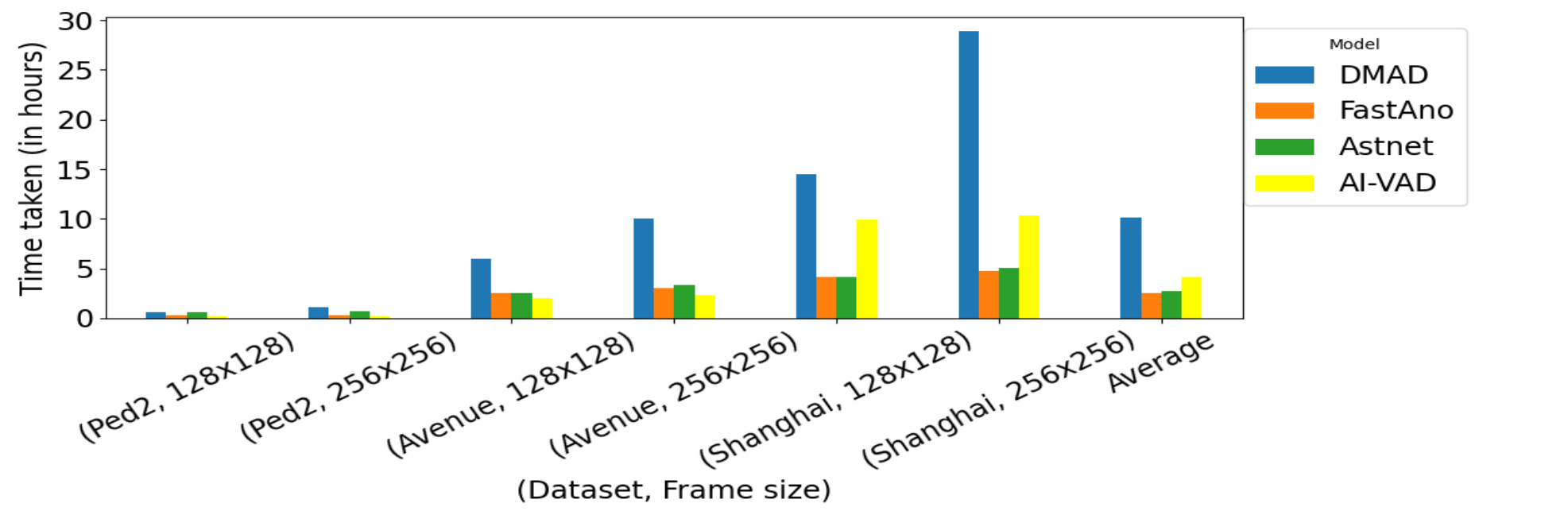

Fig 6: CPU training times of various models by dataset and frame size for 100% dataset

We observe that in Figure 6, all four models has almost had the same training time in PED2 dataset for the frame size 128x128, but when the dataset is changed there is the significant change in the CPU times. AI-VAD has the lowest training times in avenue data set and FASTANO has the lowest in shanghai dataset. On average, DMAD has the highest training time and FASTANO has the lowest.

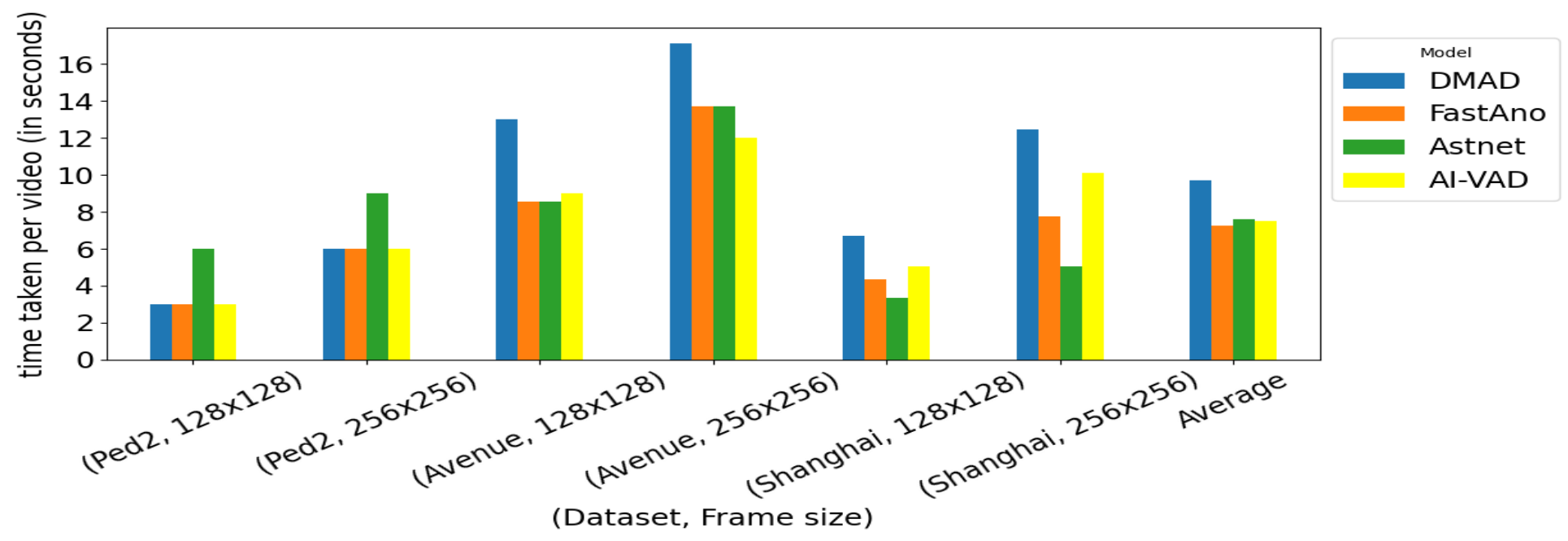

Fig 7: Inference of various models by dataset and frame size per one input

We observe that in Figure 7, on average, DMAD consumes the highest amount of time, while FASTANO takes the least amount of time. However, ASTNET does consume highest amount of time in ped2 and the least amount of time in shanghai dataset, showing its highest sensitivity to different datasets.

Conclusion

This survey evaluates the effectiveness of various anomaly detection models using publicly available datasets of varying sizes. We assess the performance of multiple machine learning algorithms in detecting anomalies in video streams, comparing a range of algorithms that employ distinct methodologies for anomaly detection. Our analysis focuses on two main aspects: a) the accuracy of these algorithms in identifying anomalous events, and b) computational efficiency, measured by the time required for both training and testing phases. Additionally, we explore the impact of data pre-processing techniques on model effectiveness. Detailed analyses of these pre-processing strategies and their outcomes are presented, particularly regarding the reduction of problem size to enhance the performance of selected VAD models. The survey also emphasizes the importance of tailoring anomaly detection systems to specific datasets and frame sizes. Overall, we found that AI-VAD is the most accurate model while only having slightly higher computational overhead compared to the fastest model. We hope our study provides insights to the model, training dataset size, and frame size selections for other VAD users.

Key References

[1] Liu, W., Chang, H., Ma, B., Shan, S., & Chen, X. (2023). Diversity-Measurable Anomaly Detection. ArXiv. /abs/2303.05047

[2] Reiss, T., & Hoshen, Y. (2022). Attribute-based Representations for Accurate and Interpretable Video Anomaly Detection. ArXiv. /abs/2212.00789

[3] Park, C., Cho, M., Lee, M., & Lee, S. (2021). FastAno: Fast Anomaly Detection via Spatio-temporal Patch Transformation. ArXiv. /abs/2106.08613

[4] Le, VT., Kim, YG. Attention-based residual autoencoder for video anomaly detection. Appl Intell 53, 3240–3254 (2023)