

Faster, Lighter, More Accurate: A Deep Learning Ensemble for Content Moderation

Authors : Arian Hosseini, Mahmudul Hasan

Abstract : To address the increasing need for efficient and accurate content moderation, we propose an efficient and lightweight deep classification ensemble structure. Our approach is based on a combination of simple visual features, designed for high-accuracy classification of violent content with low false positives. Our ensemble architecture utilizes a set of lightweight models with narrowed-down color features, and we apply it to both images and videos. We evaluated our approach using a large dataset of explosion and blast contents and compared its performance to popular deep learning models such as ResNet-50. Our evaluation results demonstrate significant improvements in prediction accuracy, while benefiting from 7.64x faster inference and lower computation cost. While our approach is tailored to explosion detection, it can be applied to other similar content moderation and violence detection use cases as well. Based on our experiments, we propose a "think small, think many" philosophy in classification scenarios. We argue that transforming a single, large, monolithic deep model into a verification-based step model ensemble of multiple small, simple, and lightweight models with narrowed-down visual features can possibly lead to predictions with higher accuracy.

Keywords : deep classification, content moderation, ensemble learning, explosion detection, video processing

Conference Title : ICCV 2024 : International Conference on Computer Vision

Conference Location : Vancouver, Canada

Conference Dates : September 26-27, 2024