AI-powered Counterfeit Products Detection Solution for Consumer Brands

Anshul Garg
Kumar Shubham
Sanket Patil
Karthik Bettadapura

Semantics Dept., Dataweave Pvt. Ltd., Singapore
Email:anshul.garg@dataweave.com
Email:shubham@dataweave.com
Email:sanket@dataweave.com
Email:karthik@dataweave.com

Abstract

With the rise in e-commerce adoption by consumers worldwide, the sale of counterfeit products on open marketplaces is an increasingly challenging problem faced by shoppers, online retailers, and consumer brands. Recent surveys have reported that over a third of shoppers have had counterfeits delivered to them on placing an order online. These defective products can have a severe impact on the brand value of retailers and consumer brands, damaging consumer trust and potentially having an adverse impact on sales volumes. If one considers the large amount of products being sold over retailer websites, the problem of identifying counterfeits only magnifies.

In this paper, we propose an efficient and robust method of detecting counterfeit products at scale using computer vision and deep learning applied on online product catalog data. The approach entails identifying small variations between two images. Essentially, the original brand manufacturer's image on its own website is compared with images listed on online marketplaces by third party sellers to detect variations, which indicate whether a product listing is legitimate or not. We use pre-trained Convolutional Neural Network (CNN) based models to take advantage of transfer learning and further fine tune them on internal data to focus on fine grained image features. In addition, we use several image processing and matching techniques based on image signatures, key points, and descriptors to achieve better accuracy.

DataWeave's Counterfeit Products Detection solution is powered by datasets that we have built in house over the years, consisting of millions of products collected from thousands of retail websites across geographies. The dataset has hierarchical information pertaining to retail taxonomy.

Detecting counterfeits on e-commerce websites where the volume of data can be in the range of terabytes is a major challenge. Our technology platform efficiently stores data pertaining to millions of processed images in Internet Archive's ARC files and maintains the indexes in lucene search engines. The entire pipeline is automated and connected via big data technologies like Kafka and orchestrated across servers using Celery.

Key Words : Counterfeit detection, CNNs, Transfer learning, Kafka, Machine Learning, Image signatures, Solr