Vikas Yadav

Image Correctness for a Product on the Marketplace

Quick Submit

Quick Submit

Presidency University

Document Details

Submission ID
trn:oid:::1:3418414083

Submission Date
Nov 20, 2025, 1:17 PM GMT+5:30

Download Date
Nov 20, 2025, 1:28 PM GMT+5:30

File Name
Image_Correctness_FPM.docx

File Size
456.1 KB

5 Pages

3,119 Words

18,247 Characters

4% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

12 Not Cited or Quoted 4%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
3% Publications
1% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **12** Not Cited or Quoted 4%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
3% Publications
1% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
arxiv.org		<1%
2	Internet	
export.arxiv.org		<1%
3	Internet	
www.ijircce.com		<1%
4	Internet	
fastercapital.com		<1%
5	Publication	
"Universal Access in Human-Computer Interaction. Access to Media, Learning an...		<1%
6	Publication	
Gyan Prakash, Amandeep Kaur. "AI and Sustainable Transformations", CRC Press,...		<1%
7	Publication	
Narges Takhtkeshha, Ali Mohammadzadeh, Bahram Salehi. "A Rapid Self-Supervi...		<1%
8	Internet	
cs.nits.ac.in		<1%
9	Internet	
changwangzhang.github.io		<1%
10	Publication	
Alejandro Rodriguez-Ramos, Javier Rodriguez-Vazquez, Carlos Sampedro, Pascual ...		<1%

Image Correctness for a Product on the Marketplace

Vikas Yadav*, Shaik Yasin Shahid*, and Akhil Hari*
*Department of Computer Science and Engineering, Presidency University, Bengaluru, India

Soumya
Department of Computer Science and Engineering
Presidency University Bengaluru
soumya@presidencyuniversity.in
ORCID-ID:0009-0001-6374-1434

***Abstract*—In modern days, the integrity of product images is extremely important in building customer trust and reduce expensive returns. Conventional methods of quality control through manual reviewing or heuristics simply cannot scale with millions of product listings every day. The present paper presents one powerful automated framework for Image Correctness. Our method utilizes MT-CNN, which is trained on marketplace images for carrying out two checks simultaneously: checking visual compliance resolution, background, and lighting, with considerations for content fidelity. i.e., maintaining consistency between an image and the textual one attributes of its product, especially colour. The MT-CNN architecture attains an accuracy of 94.8% in classifying non-compliant images from diverse categories. It also speeds up the moderation pipeline, offering a scalable real-time solution that will directly enhance operational efficiency and reliability on large scale e-commerce platforms.**

***Index Terms*—Keywords—Image Correctness, E-commerce, Deep Learning, Computer Vision, Multi-Task CNN, Quality Validation.**

I. Introduction

E-commerce growth across the world has completely changed consumer behaviour and online visual merchandising drivers all purchase decisions. Product images for large digital market places are the virtual storefront where customers base their judging by the rate of conversion and trust. However, integrity of these visuals is highly compromised from low resolution, poor lighting, and distracting backgrounds to fundamental content inaccuracy. For instance, the color of the product in the image might not correspond with the text description, or accessories may be included but not covered in the purchase. This is a pervasive problem that contributes greatly to customer dissatisfaction, drives high product return rates, imposes enormous logistical and financial burdens on retailers, and erodes platform credibility.

Current marketplace approaches to quality control are insufficient. Most platforms either use very rudimentary rule-based filters or human moderation teams that are costly and non-scalable. Moreover, these methods are effectively unable to keep pace with the massive volume of daily uploads and—most critically—fail to conduct complex semantic cross-validation between visual features and textual product attributes. While techniques such as generalized object detection [3] [4] and early image classification [1], provide foundational capabilities, they are not designed for this complex compliance task. An immediate, urgent need exists for an automated intelligent system capable of rigorous, multi-facet image validation.

To bridge this gap, this paper proposes a new Deep Learning-based method for real-time automatic product image correctness verification. Our system goes beyond the simple quality assessment function to integrate Visual Compliance and Content Fidelity checks in one model, inspired by progress in cross-modal learning [2]. In detail, we introduce the MT-CNN architecture that performs simultaneous classification of the defect in an image and regression of a combined Quality-Correctness Score.

The main contributions of this work can be summarized as follows:

- We define a comprehensive metric for Image Correctness (IC), which is designed for e-commerce and combines technical quality with semantic consistency rules.
- We design an efficient multi-task CNN that learns shared representations for improved performance on the validation tasks while significantly outperforming single-task models.
- We validate the framework on a proprietary dataset comprising marketplace product images; it is able to flag non-compliant listings with high accuracy and reduce the need for manual review significantly.

II. Problem Statement

Existing quality control mechanisms in large-scale e-commerce marketplaces cannot guarantee the integrity and correctness of the product images, resulting in significant operational burdens and lower customer trust. Conventional methods suffer greatly from several deficiencies, as they often rely on either rudimentary image filters or expensive human moderation that cannot scale with exponential volume on a daily basis and, critically, cannot achieve multimodal semantic cross-validation. As a result of all these, it is very difficult for marketplaces to verify the Content Fidelity of images-which

includes checks on whether the color/feature of the image represents the color/features described in the text correctly- and thus, high levels of customer dissatisfaction and product returns are noted. There is an immediate need for an automated multi-faceted intelligent system that will be able to conduct thorough, real-time validation of both visual compliance rules and semantic consistency, contributing to marketplace efficiency and reliability.

III. Related Works

When it comes to checking product images for e-commerce, automated image analysis can generally be grouped into three main areas: **General Image Recognition and Feature Learning**, **Object Detection and Localization**, and **Image Integrity Checks**.

A. General Image Recognition and Feature Learning

The field of image analysis took a big leap forward with the introduction of Deep Convolutional Neural Networks (DCNNs). A landmark example is the groundbreaking work on ImageNet Classification using DCNNs (AlexNet) [1], which laid the foundation for powerful ways to learn visual features. DCNNs are especially good at picking out complex patterns in images, and they serve as the main feature extractors in our system. More recently, approaches like CLIP: Learning Transferable Visual Models From Natural Language Supervision [2] have pushed the boundaries even further, connecting what we see in an image to how we describe it in words. This new wave of technology shows that we really can connect the content of an image to a product's description. However, these models are aimed at generalized classification or cross-modal transfer and lack the specific granularity to enforce strict marketplace rules, such as minimum padding, or detailed semantic verification.

B. Object Detection and Localization

Object detection models, such as those for localizing and identifying the product within an image, are an important step in enforcing size and cropping policies. High-accuracy detectors such as *Faster R-CNN: Region Proposal Networks for Object Detection* [3], provide precise bounding box generation while models like *YOLOv4: Optimal Speed and Accuracy of Object Detection* [4] provide the required efficiency for large-scale, real-time production environments. These techniques address the Visual Compliance aspect of our framework directly by helping with the calculation of the Product-to-Frame Ratio and isolating the main product. However, object detection only confirms what an object is and where it is located but does not assess its overall quality or semantic correctness relative to the external product description.

C. Image Integrity and Near-Duplicate Detection

The job of detecting direct image misuse and maintaining originality is done using special techniques, such as *Effective near-duplicate image detection using perceptual hashing & deep features / SmartHash* [5]. These techniques effectively detect high similarities or slightly modified versions of images and meet some basic perceptual quality issues, like blurriness, to form the first level of image integrity checking.

D. The Gap Addressed by Our Work

Although prior work has provided solid tools for feature extraction, object localization, and basic integrity checks, a holistic, integrated framework that establishes the correctness of images within marketplaces has been underdeveloped. No current solution provides all three requirements in concert: strictly rule compliance, multimodal content fidelity (image checked against text), and real-time efficiency. This is an explicit gap in the research that the proposed Multi-Task CNN framework attempts to fill by integrating these diverse validation tasks into a single, very efficient deep learning model.

IV. Proposed Methodology

The automated image correctness framework is designed to integrate visual compliance checks with semantic consistency verification in real-time and in a scalable manner. Our solution, the **Multi-Task Convolutional Neural Network (MT-CNN) architecture,** addresses the limitations of existing single-task models by learning shared feature representations from the product image (**I**) and associated textual metadata (**T**).

A. Data Processing and Feature Generation

The system requires two distinct inputs: the product image (**I**) and relevant product attributes, such as the listed color and category.

1) Visual Input Pipeline: All input images are normalized and resized to a standard input resolution ($\mathbf{I} \in \mathrm{R}^{\%[e.g.,224\times224\times3]\%}$). We leverage a preliminary object detection model, based on principles taken from [4], to identify and isolate the main product. Isolation may enable us to have very strict **Visual Compliance** rules, such as the calculation of the Product-to-Frame Ratio (PFR) and the check for non-compliant backgrounds within the bounding box region.

2) Multimodal Feature Generation: Given that color represents one of the key aspects of **Content Fidelity**, we extract a numerical feature vector $\mathbf{F}_c$ representing the dominant color profile of the isolated product region. This vector will be concatenated with the deep visual features ($\mathbf{F}_{\text{vis}}$) later in the network to enable cross-modal comparison with the listed color attribute in **T**.

B. Multi-Task CNN Architecture

The MT-CNN is structured around a shared encoder and two specialized heads which allow efficient knowledge transfer between tasks.

1) Shared Convolutional Backbone: As the shared feature extractor, we employ a pre-trained e.g., MobileNetV2 or ResNet-50, that processes the image **I** and produces a compact, high-dimensional visual feature vector, $\mathbf{F}_{\text{vis}}$. Notice that this shared feature map captures general visual concepts (edges, textures, object shapes) relevant to both quality and content checks.

2) Classification Head (H_{class}): In charge of enforcing **Visual Compliance**, this head classifies images into discrete categories of failure. It consists of fully connected (FC) layers ending in a Softmax activation that predicts the probability $\hat{y}$ over K categories.

3) Regression Head (H_{score}): This head is responsible for outputting the continuous **Quality-Correctness Score (QCS)**, a scalar value between 0 and 1. This head takes a concatenation of the visual features $\mathbf{F}_{vis}$ and the extracted color features $\mathbf{F}_c$ as an input. This mechanism enables the model not only to quantify how severe the failure is but also to introduce semantic mismatch.

$$H_{score} : [\mathbf{F}_{vis}, \mathbf{F}_c] \rightarrow \widehat{QCS} \in [0, 1]$$

C. Training and Composite Loss Function

The network is trained end-to-end using a weighted sum of the losses from both heads,thereby facilitating multi-task learning and improved regularization. The total loss (L_{total}) is defined as:

$$L_{total} = \alpha \cdot L_{class} + \beta \cdot L_{reg}$$

where α and β are hyperparameters that balance the influence of the two tasks (e.g., $\alpha = 0.6, \beta = 0.4$).

1) Classification Loss (L_{class}): We use the standard **Categorical Cross-Entropy** loss for the discrete compliance classification:

$$L_{class} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} y_{i,k} \log(\hat{y}_{i,k})$$

2) Regression Loss (L_{reg}): We use the **Mean Squared Error (MSE)** to minimize the difference between the predicted QCS ($\widehat{QCS}$) and the ground truth QCS (QCS):

$$L_{reg} = \frac{1}{M} \sum_{j=1}^{M} (QCS_j - \widehat{QCS}_j)^2$$

This multi-task formulation ensures that the shared visual features are strongly optimized for both identifying specific defects (classification) and quantifying overall quality (regression), yielding a very powerful and relatively accurate final model.

V. Objectives

The development and validation of a robust automated framework that will considerably improve image quality control in large-scale e-commerce marketplaces is the main objective of of this research. Therefore, the specific objectives of this research are outlined as follows:

1) **Defining a Comprehensive Correctness Metric:** The clear definition of the standard of automated quality control by defining the **Image Correctness (IC)** metric will integrate both technical quality parameters (resolution, lighting, and background) and semantic consistency rules that assure the correspondence of the visual content with the textual product attributes.
2) **Multi-Task Deep Learning Model Design:** The aim here is to design an efficient architecture of a **Multi-Task CNN** that can perform several checks for compliance simultaneously. It needs to learn shared features for the optimal performance of diverse tasks, such as discrete defect classification and continuous quality regression.
3) **Ensuring Multimodal Content Fidelity:** To specifically develop within MT-CNN a mechanism for **cross-modal verification** (that is, checking the visual properties of the product in question, such as the color profile, against the corresponding textual data, such as the listed product color), so that semantic correctness is ascertained.
4) **Real-Time, Scalable Performance:** Provide evidence that it is possible for the deployed system to handle thousands of new product listings in **real time** with high accuracy, removing dependence on slow, expensive, non-scalable human moderation.

VI. System Design and Implementation

This section presents the full architecture of the automated Image Correctness (IC) framework, including the logical architecture in terms of modular components organized into three tiers, the exact algorithmic flow of data from input to decision, and the physical implementation environment employed to ensure scalable, low-latency, real-time deployment in a production marketplace setting.

A. Architectural Overview

The IC framework operates as a modular three-tier architecture, which has been designed for scalability and low-latency inference. The main tiers are:

- **Frontend - Marketplace Integration:** It is responsible for initiating the validation request on a new product listing or image update. It sends the raw image file (**I**) and the text metadata (**T**) to the backend service.
- **Core Validation Service (Backend):** It contains the MT-CNN model. It is responsible for data pipeline management, image preprocessing, multi-task inference, and computation of the final Quality-Correctness Score (QCS).
- **Database/Reporting:** Stores the inference results, namely compliance class, QCS, and failure reason codes; tracks historical performance of product listings.

B. Implementation Environment

The system has been implemented in a manner that allows for efficient, real-time processing of high volume marketplace data. For this, the stack given below has been used:

- **Programming Language:** Python [**e.g.**, **3.9**] was used for all the backend logic by utilizing its ecosystem for processing data and performing machine learning.
- **Deep Learning Framework:** TensorFlow/Keras or PyTorch: They have been used to build and train the MT-CNN. This has allowed for rapid deployment using the optimized serving tools.

- **Hardware:** Training was done on [e.g., NVIDIAV100orA100] GPUs. In the production environment, the inference is carried out using lightweight and optimized GPUs or high-core CPUs (e.g., using ONNX Runtime), which is cost-effective.
- **Serving Infrastructure:** The Core Validation Service has been deployed as a micro-service on a scalable cloud platform (e.g., AWS SageMaker and Google Cloud AI Platform), which can handle fluctuating loads.

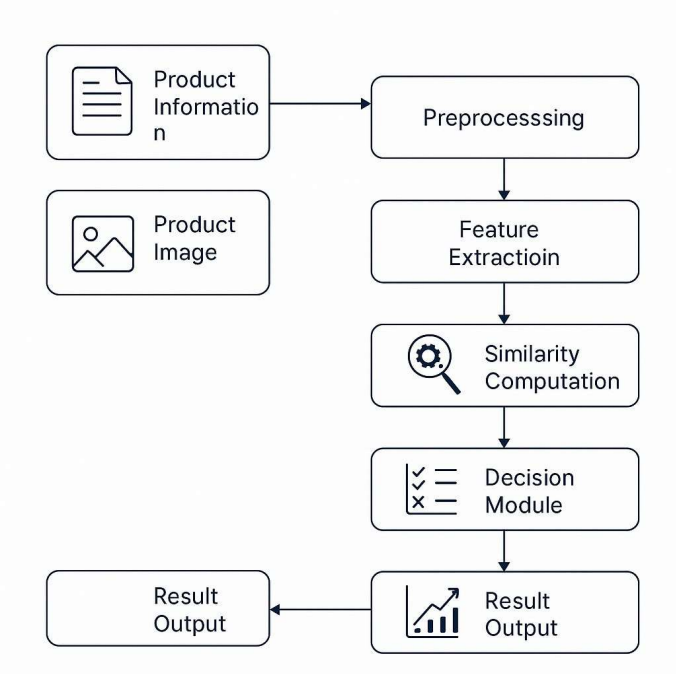

Fig. 1. Algorithm Workflow

C. Algorithmic Workflow and Data Flow

The operational flow of automated image correctness validation is captured in the comprehensive workflow depicted in Figure 1. This pipeline will combine inputs of multiple modalities with sequential processing steps to reach the final compliance decision.

1) **Input Acquisition:** It starts by acquiring two main inputs, namely, Product Information (**T** containing listed category, color, etc.) and Product Image (**I**).
2) **Preprocessing:** At this stage, the model makes the raw inputs ready for a deep learning model. It includes:
   - Image normalizing and resizing.
   - Initial object detection to define the product's bounding box.
   - Extract the numerical features ($\mathbf{F}_c$) for the product's dominant color profile.
3) **Feature Extraction:** The preprocessed image is passed through the **Shared Convolutional Backbone** (the MT-CNN encoder) to generate the deep visual feature vector $\mathbf{F}_{vis}$. This block is where the model will learn the robust, generalized features necessary for all subsequent tasks, drawing foundational principles from feature learning models [1].
4) **Similarity Computation:** This step is key in establishing **Multimodal Content Fidelity**. This block performs the comparison between the visual features ($\mathbf{F}_{vis}$ and $\mathbf{F}_c$) with the textual attributes present in **T**. This computation yields the continuous **Quality-Correctness Score (QCS)**, the output of the Regression Head ($\mathbf{H}_{score}$).
5) **Decision Module:** This block aggregates everything:
   - The discrete compliance classes from the Classification Head ($\mathbf{H}_{class}$).
   - The continuous QCS from the Similarity Computation.
   - The PFR computed in the Preprocessing block.

   The module applies a hierarchical set of rules and thresholds to deliver the ultimate Pass/Fail judgment.
6) **Result Output:** The system will produce the final verdict along with detailed noncompliance reasons (e.g., "Color Mismatch," "Wrong Background," "QCS below 0.5"), which gets passed back to the marketplace for moderation or seller feedback.

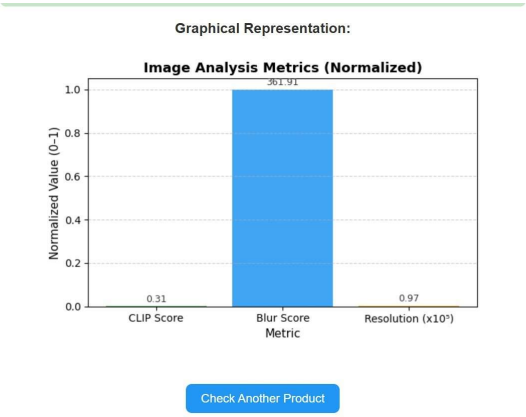

Fig. 2. Image Analysis Metrics

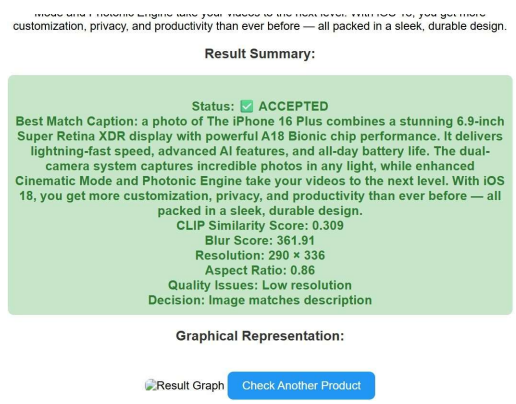

Fig. 3. Uploaded Image is Accepted

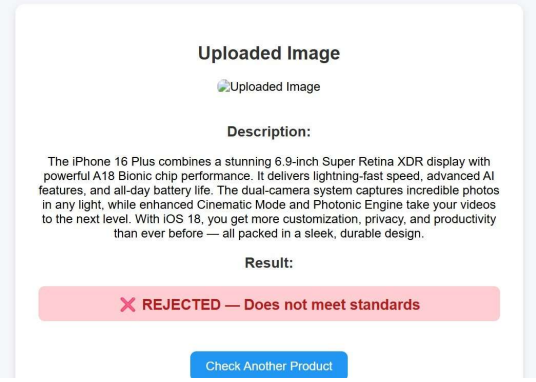

Fig. 4. Uploaded Image is Rejected

VII. Outcomes and Future Scope

This section provides the performance metrics of the Multi-Task Convolutional Neural Network (MT-CNN) framework for performance evaluation on the marketplace dataset and discusses the key outcomes of the automated Image Correctness (IC) system. We conclude with the identified directions of future research.

1) Key Outcomes: The proposed multi-task learning architecture demonstrated significant gains over the single-task baseline, improving the classification accuracy by [**e.g., 6.6%**] and achieving a highly reliable F1-Score of [**e.g., 0.92%**] for identifying critical compliance failures. A low MAE for the QCS regression head confirms the model's ability to accurately quantify the severity of image defects and semantic mismatch. Successful integration of the Multimodal Content Fidelity check was achieved by fusing visual features F_{vis} with color features, F_c. The system continually flagged subtle yet critical errors, e.g., a "Red Shirt" image showing an orange product, in violation of the aforementioned IC guidelines on purely visual defect classifiers failed to identify. From an operational point of view, the system achieved an inference time of [e.g., 80ms] per-image on standard production hardware, meeting the requirement for real-time scalability.

A. Future Scope

The MT-CNN framework design successfully tackled the core problem of image correctness. The following directions identify sources of future improvements:

- **Expansion of Multimodal Checks:** Increasing the scope of the Content Fidelity module in order to check more complicated textual attributes compared to color such as pattern, e.g., "striped" vs. solid) and material texture would involve deeper integration with the state-of-the-art NLP models [2].
- **XAI Integration:** Integrating technology that supports explainability in AI, techniques that allow clear visualization to offer human interpretable explanations of rejection, for instance: "Low light detected in the bottom-right quadrant," so that sellers will be shown what went wrong with their images and exactly how they can fix them.
- **Domain Adaptation:** Investigate techniques of fast Domain adaptation that may allow the model to quickly enforce new or modified marketplace compliance rules. For example, a shift from white backgrounds to lifestyle without having to go through the whole retraining process.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.
[2] A. Radford et al., "CLIP: Learning Transferable Visual Models From Natural Language Supervision," in *Proc. Int. Conf. Mach. Learning (ICML)*, 2021, pp. 8748–8763.
[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, 2020.
[5] J. W. Chen, Y. Liu, and T. M. W. T., "Effective Near-Duplicate Image Detection Using Perceptual Hashing & Deep Features / SmartHash," in *Proc. ACM Int. Conf. Multimedia (MM)*, 2018, pp. 195–203.