

A Multiple-Loss Dual-Output Convolutional Neural Network for Fashion Class Classification

Okeke Stephen*, Uchenna Joseph Maduh**, Sanjar Ibrokhimov*, Kueh Lee Hui***, Ahmed Abdulhakim Al-Absi****and Mangal Sain*****

*Department of Ubiquitous Information Technology, Dongseo University, Busan, South Korea

**Department of Civil Engineering, Faculty of Engineering, Yeungnam University, Gyeongsan, South Korea

***Dept. of Electrical Engineering, Dong-A University, South Korea

****Department of Computer Engineering, Kyungdong University- Global Campus, Gangwondo, South Korea

*****Division of Information Engineering, Dongseo University, Busan, South Korea

okeke2020@yahoo.com, ujmaduh@gmail.com, ib.sanjar93@gmail.com, leehkueh@dau.ac.kr, absiahmed@kduniv.ac.kr and mangalsain1@gmail.com

Corresponding Author: leehkueh@dau.ac.kr

Abstract— An improved multi-loss multi-output convolutional neural network method was deployed to extract features from a given set of disjointed data (Fashion and Color) with diverse convolutional chunks in a single network. The first convolution block extracts features from the first image dataset (Fashion) and determines the classes to which they belong. The second block is responsible for learning the information encoded in the second set of data (color), classify and append such to the features extracted from the first convolutional block. Each block possesses its loss function which makes the network a multi-loss convolutional neural network. A set of double fully connected output heads are generated at the network terminal; enabling the network to perform predictions on a combination of disjointed labels. To validate the classification ability of our network model, we conducted several experiments with different network parameters and variations of data sizes and obtained remarkable classification results of 98 and 95 on the fashion and color sets respectively.

Keywords— Multi Loss Multi Output Convolution, Convolutional Neural Network, Fashion Class Classification

I. INTRODUCTION

In recent time, object detection, recognition and segmentation [1, 2] have recorded remarkable performance in computer vision tasks. This breakthrough [16] substantially improved the handcrafted method of object recognition and detection systems. This pioneered several researches in this area and successive enhancements in various problems relating to fashion; like cloth classifications, prediction of several attributes of a distinct bit of clothing and enhancing clothing image retrievals [3, 4]. Prominent e-commerce lines are increasing their commitment and investment in automated fashion. Better knowledge of users'

unique preference and recommending same to them has become a crucial task.

Lately, many models such as recommendation and personalization based algorithms have seen a ton of research interests; including techniques driven by collaborative filtering [5], recommendation-driven content systems like neural networks, probabilistic graph models and a combination of these systems - hybrids. Collaborative filtering [5] works by analysing people's behaviour and favourite items thereby linking them to a pattern that is predefined in order to recommend a product to them. However, attributes or features are deployed by content-based techniques to recommend items to users relying solely on users search habits. On the other hand, knowledge-based systems [5] (hybrid system) integrate product features and user preference to recommend a product for them.

In addition, texture-based synthesis attempts to understand the fundamental texture of an image so as to create new samples that possess identical texture. This research area is solely fixated on parametric and non-parametric techniques [7]. Resampling of particular pixels is employed by non-parametric method in an attempt to form the image or deploy particular patches from the original and create a new image [8][9]. Statistical models are used to define parametric methods that denote the texture [10]. A convolutional neural network based parametric model was developed in 2015 [7] for texture synthesis. Extraction of feature maps

was employed to model the style of a specific image by feeding the image into a pre-trained CNN model by employing a VGGNet with 19 layers. They effectively split the content and style of a randomly selected image and establish the way in which styles can be embedded in them by applying prior textures. In this work, we introduce regularization technique to enhance the existing multiple-loss multiple-output CNN based system to classify fashion images and learn their respective colors from different network blocks.

II. RELATED WORKS

Detail overviews of some existing fashion classification and segmentation methods have been explored over the years. A method that deal with clothing parsing using deep learning models was tested successfully [11]. They generated Fashionista dataset containing 158,235 images and from this dataset 685 images were deployed for training and testing purposes. Image retrieval approach was employed in [12] to handle cloth parsing task. Image retrieval method was extensively explored in [13, 14], garment and style classification with fine-grained technique was examined in [15, 16], landmark detection in fashion [17] and domain transfer with image conditioning in [18]. Also studied are subjective characteristics that deal with fissionability [19], popularity [20] and several approaches that rely on collecting voluminous image datasets by crawling the internet. Weak labels [21] were also examined as were annotations that use manual method on image through crowd-sourcing [22].

Other recent works have also demonstrated the use of retrieval methods that adopt cross-scenario system [23, 24], street-to-shop [24] technique that solves the discrepancy in domain among photos in streets and photos in shops applying sparse representations. Methods like image patch-match and object retrieval based on fine grain system or face verification rely on similarity learning capability of deep neural networks. These systems learn representations in conjunction with either their generic learned multi-layer similarity network measures or distance functions. Our approach is narrowly related to multi-task convolutional neural network method but utilizes individual network

section for learning distinct classification tasks designated to it.

III. METHOD

Our model consists of two disjointed convolution blocks with separate loss functions. The first block B1 as shown in Figure 1, is responsible for learning and extracting features from the fashion images and then classifying them into the categories in which they belong to. It consists of 5-layer sub-blocks with 100x100 image input and the first layer having (32x3x3) feature map. The remaining layers compose of (64x3x3), (64x3x3), (128x3x3), (128x3x3) and (256x3x3) sizes respectively with each sub-layer block having dropout, pooling, batch normalization, padding and regularization functions. The second major block B2 deals with learning the color information in the color dataset. It is composed of 4 smaller layer sub-blocks with (16x3x3), (32x3x3), (32x3x3) and (128x3x3) layer sizes respectively. Each smaller sub-layer blocks containing dropout, pooling, batch normalization, padding and regularization terms. In summary, the network has 5,605,255 total parameters, 5,603,495 trainable parameters and 1,760 non-trainable parameters. We introduced L2 regularization term in each layer which adds extra penalty terms to some parameters in the network. The size of the input image remains unaltered throughout the training process in the B1 block.

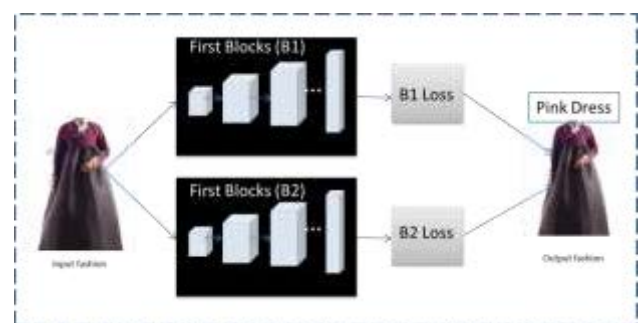


Figure 1. Multiple Loss Multiple Output CNN Architecture

IV. EXPERIMENT ANALYSIS

A. Dataset

To demonstrate the validation accuracy of our proposed model, we used the dataset provided by [27]. The dataset consists of 2,167 fashion images of 6 categories including: blue dresses having 386

images, black jeans containing 344 images, blue shirts composing of 369 images, red dresses of 380 images, red shirts of 332 images and blue jeans consisting of 356 images. 80 percent of the images were allocated to training set and the remaining 20 to validation set.

B. Training Process

To evaluate our model, we used different sizes of the training images, variant training samples, batch of different sizes, several epochs and alterations of filter kernels from the dataset. Also, to eliminate overfitting, we introduced L2 regularization terms and smoothened the learning process by applying learning rate scheduler starting from $1e-4$. All the experiments were performed using Keras and Tensorflow open source deep learning libraries in conjunction with NVIDIA's CUDA 9.0 and cuDNN library 7.0. Training was conducted on a desktop with NVIDIA TitanX GPU and 12GB memory. To mitigate random influence, all experiments were repeated 10 times. The filter kernels in the network were initialized via sampling emanating from usual standard network distribution and the feature map biases were all initially set to zero. Model optimization was done using Ada optimizer across the entire training dataset having the following configurations: learning rate was initially set to $1e-4$ and annealed with $\lambda \times 1e-3 * 0.9^{**} x$, over every 100 epochs and L2-norm weight decay in all the filter kernels was fixed to 0.01.

C. Results

A lot of technique like dropout, regularization and batch normalization were employed to fit our small dataset into our large model. As shown in figure 2, the increase in epoch size and samples of training set, the increase in both training and test accuracy we obtain. We solely used 100 epochs in the entire experiment because the marginal utility of epochs reduced at approximately 100 epochs. In addition, we adopted smaller batch size to improve the classification performance of our model although with higher computation demand and time.

In figure 2, the top plot represents the training and test accuracies we obtained in the fashion network block and the second plot illustrates the output of the color network block respectively. On both network blocks, we obtained 100% and 99.9%

training accuracy on both the fashion and color dataset respectively, 98% and 95% test accuracy with training size of 100×100 on both dataset.

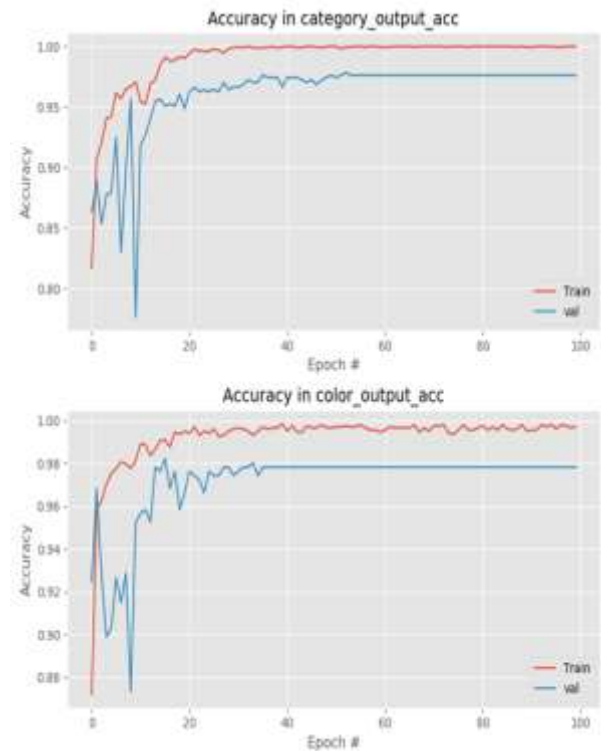


Figure 2. Training and Test Accuracies on the Fashion and Color Dataset

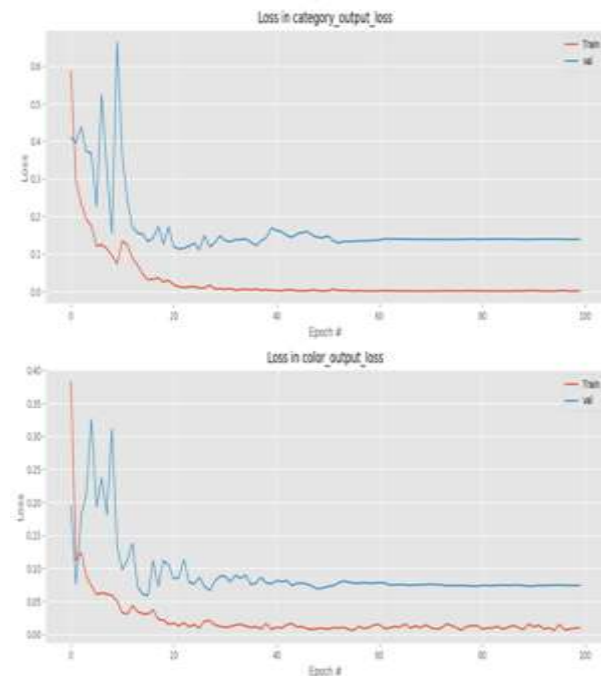


Figure 3. Training and Test Loss on the Fashion and Color Dataset

Figure 3 illustrates the training and test losses obtained from both the fashion and color datasets respectively and Figure 4 shows the overall loss obtained from both dataset. We obtained 0.094 fashion loss, 0.107 color loss and total loss of 0.0043 as shown above. We used moderate batch-size to smoothen the iteration so as to make optimal use of the GPU. Overall, our model achieved 98% fashion set classification accuracy and 96% color set classification accuracy.

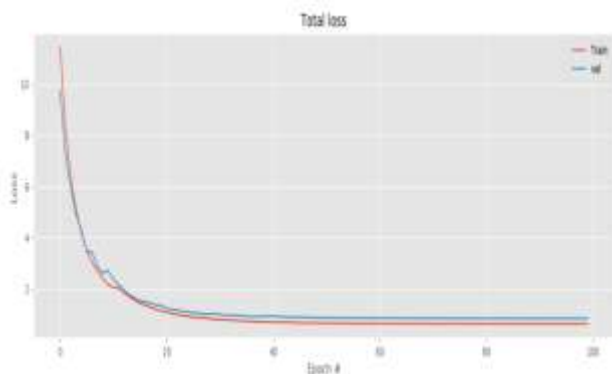


Figure 4. Total Training and Validation Loss on the Fashion and Color Dataset

V. CONCLUSIONS

We presented a multiple-loss multiple-output convolutional network for joint classification of fashion categories and their associated color information. By ushering in regularization, normalization and learning rate annealing methods. We trained a more efficient disjointed network to learn from two disjointed datasets and perform classifications on them. We demonstrated how classification and predictions can be learned concurrently to generate an enhanced performance for the two tasks. In future, we plan to introduce more dataset and classification scenarios to improve the accuracy of the proposed model.

ACKNOWLEDGMENT

REFERENCES

- [1] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Adv in Neural Inf Process Syst* 2012; 1097-1105.
- [2] Ren S, He K, Girshick RB, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR* 2015; abs/1506.01497.
- [3] Liu Z, Luo P, Qiu S, Wang X, Tang X. Deep fashion: Powering robust clothes recognition and retrieval with rich annotations. In: *Proc of the IEEE Conf on Comput Vision & Pattern Recog* 2016; pp. 1096-1104.
- [4] Kalantidis Y, Kennedy L, Li LJ. Getting the look: clothing recognition and segmentation for automatic productsuggestions in everyday photos. In: *Proc of the 3rd Assoc for Computing Machinery Intconf on multimedia retrieval* 2013; pp. 105-112.
- [5] Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internetcomputing* 2003; 7: 76-80.
- [6] Trewin S. Knowledge-based recommender systems. *Encyclopedia of library and information science* 2000; 69(Sup21plement 32):180.
- [7] Gatys L, Ecker AS, Bethge M. Texture synthesis using convolutional neural networks. In: *Adv in Neural InformProcess Syst* 2015; 262-270.
- [8] Wei LY, Levoy M. Fast texture synthesis using tree-structured vector quantization. In: *Proc of the 27th annualconference on Computer graphics & interactive techniques*; ACM Press/Addison-Wesley Publishing Co. 2000; pp. 479-488.
- [9] Kwatra V, Schodl A, Essa I, Turk G, Bobick A. Graphcut textures: image and video synthesis using graph cuts. In: *Association for Comp Machinery Trans on Graphics* 2003; 22: 277-286.
- [10] Julesz B. Visual pattern discrimination. *IRE Trans on Inf Theory* 1962; 8:84-92.
- [11] Yamaguchi K, Kiapour MH, Ortiz LE, Berg TL. Parsing clothing in fashion photographs. *Proc. Computer Vision & Pattern Recognition* 2012; pp. 3570-3577.
- [12] Yamaguchi K, Kiapour MH, Berg TL. Paper doll parsing: retrieving similar styles to parse clothing items. In: *Proc.of the IEEE International Conference on Computer Vision* 2013; pp. 3519-3526.
- [13] Huang J, Feris RS, Chen Q, Yan S. Cross-domain image retrieval with a dual attribute-aware ranking network. In: *IntConf on Comput Vision* 2015.
- [14] HadiKiapour M, Han X, Lazebnik S, Berg AC, Berg TL. Where to buy it: Matching street clothing photos in online shops. In: *IntConf on Comput Vision* 2015.
- [15] Bossard L, Dantone M, Leistner C, Wengert C, Quack T, Van Gool L. Apparel classification with style. In: *Asian Conf on Comput Vision* 2012.
- [16] Chen H, Gallagher A, Girod B. Describing clothing by semantic attributes. *Enterprise Cloud Comp & Virtualization* 2012.
- [17] Yoo D, Kim N, Park S, Paek AS, Kweon IS. Pixel-level domain transfer. In: *Enterprise Cloud Computing & Virtualization* 2016.
- [18] Simo-Serra E, Fidler S, Moreno-Noguer F, Urtasun R. Neuroaesthetics in fashion: Modeling the perception of fashionability. In: *Comput Vision & Pattern Recog* 2015.
- [19] Vittayakorn S, Yamaguchi K, Berg AC, Berg TL. Runway to realway: Visual analysis of fashion. In: *Workshop on Application of Computer Vision* 2015.
- [20] Yamaguchi K, HadiKiapour M, Berg TL. Paper doll parsing: Retrieving similar styles to parse clothing items. In: *IntConf on Comput Vision* 2013.
- [21] Liu Z, Luo P, Qiu S, Wang X, Tang X. Deep fashion: Powering robust clothes recognition and retrieval with rich annotations. In: *Computer Vision & Pattern Recognition* 2016.
- [22] Liu S, Feng J, Song Z, Zhang T, Lu H, Xu C, Yan S. Hi magic closet, tell me what to wear! In: *Association for ComputingMachinery annual Conference on Multimedia* 2012.
- [23] Liu S, Song Z, Liu G, Xu C, Lu H, Yan S. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. In: *Computer Vision & Pattern Recognition* 2012.
- [24] S. Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks. In: *Computer Vision & Pattern Recognition* 2015.
- [25] Schro_F, Kalenichenko D, Philbin J. Facenet: A united embedding for face recognition and clustering facenet. In: *Computer Vision & Pattern Recognition* 2015.
- [26] Taigman Y, Yang M, Ranzato M, Wolf L. Deepface: Closing the gap to human-level performance in face verification. In: *Computer Vision & Pattern Recognition* 2014.
- [27] Ziwei L, Ping L, Shi Q, Xiaogang W and Xiaoou T. DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June, 2016.



Kueh Lee Hui is working as an assistant professor at the department of Electrical Engineering, Dong-A University since 2012. She completed her PhD Degrees from Department of Electrical Engineering, DongA University, Korea. In 2009 she completed her BS degree in Electronic and Communication, Department of Electronic Engineering,

University Malaysia of Sarawak, Malaysia. She also done MS in 2007 from Malaysia. Her research interests are image processing, face recognition, digital image forensic, intelligent control and control application, power system.



Ahmed Abdulhakim Al-Absi is an assistant professor and head of smart computing department at Kyungdong University - Global Campus in South Korea. He earned his Ph.D. degree in ubiquitous computing at Dongseo University- South Korea in 2016. His research interests include Database Systems, Big Data, Hadoop, Cloud computing, Distributed systems, Parallel computing, High-performance computing, VANET, and bioinformatics. He received a Master of Science (MS) degree in information technology at University Utara Malaysia-Malaysia in 2011, and a Bachelor of Science (BS) degree in computer applications at Bangalore University- India in 2008



Mangal Sain received the M.Sc. degree in computer application from India in 2003 and the Ph.D. degree in computer science in 2011. Since 2012, he has been an Assistant Professor with the Department of Computer Engineering, Dongseo University, South Korea. His research interest includes wireless sensor network, cloud computing, Internet of Things, embedded systems, and middleware. He has authored over 50 international publications including journals and international conferences. He is a member of TIIS and a TPC member of more than ten international conferences.