Enabling Hyper-Personalisation: Automated Ad Creative Generation and Ranking for Fashion e-Commerce

Sreekanth Vempati
Myntra Designs
Bengaluru, India
sreekanth.vempati@myntra.com

Korah T Malayil
Myntra Designs
Bengaluru, India
korah.malayil@myntra.com

Sruthi V*
Microsoft
Bengaluru, India
vsruthi98@gmail.com

Sandeep R
Myntra Designs
Bengaluru, India
sandeep.r@myntra.com

ABSTRACT

Homepage is the first touch point in the customer's journey and is one of the prominent channels of revenue for many e-commerce companies. A user's attention is mostly captured by homepage banner images (also called Ads/Creatives). The set of banners shown and their design, influence the customer's interest and plays a key role in optimizing the click through rates of the banners. Presently, massive and repetitive effort is put in, to manually create aesthetically pleasing banner images. Due to the large amount of time and effort involved in this process, only a small set of banners are made live at any point. This reduces the number of banners created as well as the degree of personalization that can be achieved. This paper thus presents a method to generate creatives automatically on a large scale in a short duration. The availability of diverse banners generated helps in improving personalization as they can cater to the taste of larger audience. The focus of our paper is on generating wide variety of homepage banners that can be made as an input for user level personalization engine. Following are the main contributions of this paper: 1) We introduce and explain the need for large scale banner generation for e-commerce 2) We present on how we utilize existing deep learning based detectors which can automatically annotate the required objects/tags from the image. 3) We also propose a Genetic Algorithm based method to generate an optimal banner layout for the given image content, input components and other design constraints. 4) Further, to aid the process of picking the right set of banners, we designed a ranking method and evaluated multiple models. All our experiments have been performed on data from Myntra (http://www.myntra.com), one of the top fashion e-commerce players in India.

1 INTRODUCTION

In the current e-commerce era, content on the homepage plays an important role in a customer's journey for most of the e-commerce companies. Significant amount of the homepage on online shopping apps/websites is dedicated to banners. Example homepage banners are shown in Figure 1. These banners play a key role in visual communication of various messages to customers such as sale events, brand promotion, product categories and new product launches. The primary components of a banner are the image, also called a creative, and the landing page associated with it.

*Work done while at Myntra

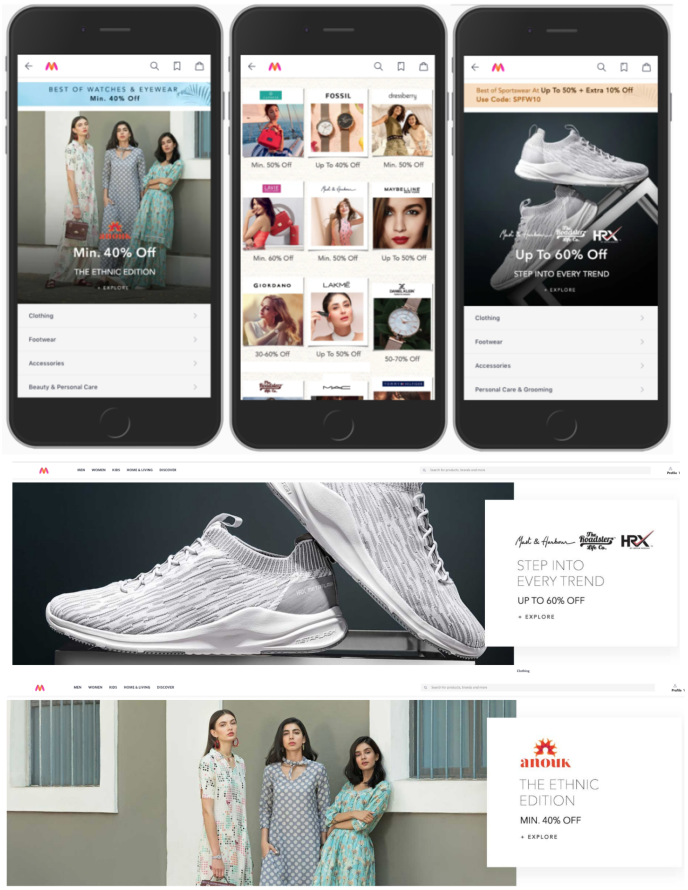

Figure 1: First row shows examples of Homepage banner/ad images on the Myntra app. Second row shows how same banner image can appear in different formats on desktop.

Any banner image is an amalgamation of an underlying theme, a background image, the brand logo in focus and affiliated text phrases. The text is further constituent of font colours, size and typography. Banners essentially come down to being one particular permutation of these components which is visually appealing and can deliver the message. Given a requirement, a designer typically

takes these various elements from a library, places these elements as per aesthetics and applies few image transformations to create a banner image.

In this paper, we show how we can automate the process that designers follow, using a set of deep learning and image processing techniques. Usually designers pick images from a large image collection called photo-shoot images. Most of these images are provided by brands or other associated entities. These images are also used on other platforms like public hoardings, offline stores etc. Designers also use product catalogue images on few occasions. These photo-shoot images are taken in good lighting conditions showcasing highlights of the brands.

Due to the notably large amount of manual hours invested in creating these banners, only a small set is made live at any point of time. This reduces both the variety of output produced and the degree of personalization that can be achieved. But if the range of available banners increases, satisfying the personal taste of a larger audience can be achieved. From a wider spectrum of options, we can now accomplish targeted banners that cater to different sectors of customers, instead of generic ones made for all. In this paper, we do not focus on the user personalization engine, which is already internally built, but we focus on large scale generation of the inputs to this engine.

We present a method which generates banner images using a library of design elements. Examples of design elements include background content image, text phrases, logo etc., One of the key design elements is the underlying layout of the banner which determines the spatial arrangement of various design elements like text, logo etc., A layout can be used to describe a group of banners, and can also be used to create new banners with the same blueprint. We use a Genetic algorithm based method which can generate a banner layout given a background image, logo and text. It takes into account, design considerations like the symmetry, overlap between elements, distance between elements etc for assessing layout quality. Overall, we can generate the banners using a library of photoshoot images and tag lines for any given themes. As input, we make use of photo-shoot images that are available to the designers in e-commerce companies. We further present an offline process to rank these generated banners. This method constitutes of machine learning models built on banner meta-data.

The use cases of this proposed method are not restricted to e-commerce websites only. It can further be extended to automating banners for social messaging platforms and movies or online video content providers.

In the next sections, we discuss some of the related work to this problem and then we talk about the method for creating the banners, touch upon few challenges and solutions. Further, we present and evaluate the methods for evaluating the banners generated using different design elements.

2 RELATED WORK

As this is an attempt to solve a new problem, there is very limited related work available for the same. We do have some research being done to solve similar problems in other domains. In the past, there have been papers which solve the problem of automated layout design for photo-book[18] and magazine cover [11].

Photo-book design system [18] was built to automatically generate photo compositions such as collages or photo-books using design and layout principles. In their paper, images and text elements are pre-processed and then content is distributed across pages using pre-defined rules. Then, a layout is generated using a set of design principles. Further, genetic algorithm is used to evaluate the layout, whose fitness function takes into account the borders, text elements and overall visual balance of the page.

There is some work on generating new layouts and also transferring the layout using an example [14][15][21]. In [14], Generative Adversarial Networks are used to generate new layouts for UX design and clip-art generation. The generator takes randomly placed 2D elements and produce a realistic layout. These elements are represented as class probabilities and bounding boxes. A CNN based discriminator is used as the aim was to be similar to the human eye and spatial patterns could be extracted.

Alibaba's Luban [10][20] is the most relevant one to our work, but we do not have any published work regarding this. As per the blog, the design team established a library of design elements and allowed Luban to extract and cluster features from raw images. Then, given a blank canvas, it places elements on it in a random fashion and then uses reinforcement learning to learn the good designs.

[21] tackles the problem of multi-size and multi-style designs, i.e. modifying a single design to suit multiple styles and sizes. Automation of layout design by optimizing an energy function based the fitness of a layout style which measures factors such as margin, relative position, etc. In [15], a new banner is being generated based on a given examples using energy function optimization.

Another energy based model is built by [16] targeting single page graphic designs. The images are analyzed to get hidden variables such as importance of different elements and alignment. The energy based model is then based on positions of different elements, balance, white space, overlap area, etc. Design re-targeting is also presented, i.e, transferring the same design to a different aspect ratio. We have adopted some of the energy terms in generating the layout in our work.

In the movies domain, Netflix[1] blogs talk about generating art work which is personalized to the user. The focus here is primarily on finding the best thumbnail for a given movie and user by using reinforcement learning approaches.

For predicting the Click-Through-Rate (CTR) of online advertisements, [6] has trained various models on a large dataset of images. In Deep CTR system [4], a deep neural network is proposed in which uses convolution layers to automatically extract representative visual features from images, and nonlinear CTR features are then learned from visual features and other contextual features by using fully-connected layers. There is also work on finding the quality of native ads using image features [23].

As our aim is to automatically generate a large number of visually appealing banners from a library of photoshoot images, brand logos and associated text, we have adopted a multi-pronged approach in solving various aspects of the problem. The original images are annotated and re-sized. The best possible layout is generated and post processing steps are performed. The final generated creatives are ranked according to their aesthetic appeal and historical performance. The entire approach is explained further.

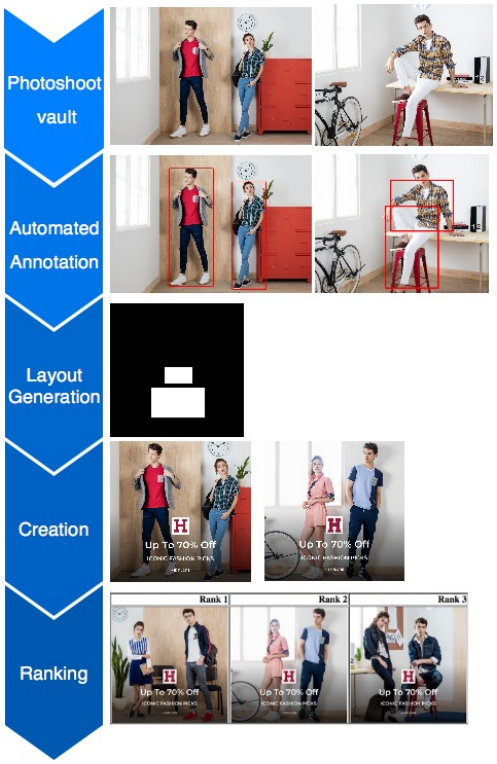

Figure 2: End-to-end pipeline for automated creation of banners

3 METHODOLOGY

Given that the integral constituents of a banner are the image, logo, text and additional elements, structuring them optimally results in the end product. This structure is defined by a layout, which is essentially the positions and sizes of the various components. Hence the input to automated generation would primarily be a layout and also the image in focus, the brand's logo, text and other associated details. We could use both human provided layout as well as generate a layout based on the given photoshoot image and design elements. We present a layout generation method in the further sections. Different layouts are generated and the best ones are utilized for final creatives generation.

For automation to be possible, we need automated tags to pick the right image for the right theme/purpose. A bare photo-shoot image is composed of various fashion articles spread across it. Examples of fashion articles are "Men-Tshirts", "Women-Jeans" etc., Information regarding this is required for filtering and selecting appropriate ones. Thus the first step would be to perform large scale automated annotation of all images and tag each of them with relevant data.

Once annotation is complete, this newly obtained information is given as input to the layout generation module and further to creation module. The region of interest is extracted, and the different elements are stitched together according to the layout specifications.

The final result produced is a large number of banners for the given theme. Since only a few from this pool of options would actually be required , these are further re-ranked by a model built on historical data.

End to end steps involved in the banner creation and ranking can be found in the Figure 2.

3.1 Automated annotation of Photo-shoot images

Automated annotation involves extracting the meta-data of an image. Simpler attributes like brand name and season name are given as a label. For all the other attributes, we need to visual understanding of the image which is done by using a set of detectors for each attribute. The constituents of the images are tagged based on categories or bounding boxes. A bounding box is a rectangle that completely encloses the object in focus and is described by coordinate points.

The different aspects of annotation ranges from the significant objects such as people or the main article, along with their types to the secondary level details such as a person's gender, the type of scenery of the image, and the number of articles present in each category. We explain more details for each of the aspects below.

3.1.1 ***Object and Person Detection.*** The various objects present in the image was detected using the MaskRCNN[9] object detector.

MaskRCNN was built for the purpose of instance segmentation, which is essentially the task of detecting and delineating distinct objects of interest in an image. The state of the art instance segmentation techniques are R-CNN[8], Fast-RCNN[17], Faster-RCNN[17] and MaskRCNN[9]. MaskRCNN was chosen as it addresses the issues of Faster-RCNN and also provides pixel-to-pixel alignment.

In particular, detecting people is of more interest to us as most of the times the article in focus is on or near them. Thus images were tagged with the bounding boxes of people present and additional information such as total number of people, dominant person, etc. We have used pre-trained detector for our person/object annotations.

3.1.2 ***Fashion Category Detection.*** Though detecting the people in image narrows down the region of interest, the actual focus is always on the product that the banner was originally meant to be created for. Tagging this product will help us give significant importance to it. Thus, to identify the various fashion categories present in an image, an in-house detector was used. This fashion category detector was built using Mask RCNN architecture[9] and was trained on fashion specific categories like shoes, watches, etc., The training data for this detector contained manually labelled bounding boxes and class labels for the fashion categories. Example detections are illustrated in Figure 3. The mean average precision of this detector (mAP) is 67.9%. This value is inline with the mAP obtained for the famous Pascal VOC 2007 dataset which is 63.4%.

This detector provides a bounding box of the category along with it's type. The types of categories include top-wear, shoes, watches, etc. Additionally, the image was also tagged with the dominant article present in terms of area. This would later be useful in filtering images based on the articles present.

3.1.3 ***Gender and Face Detection.*** Apart from the entire person's body, detecting the face will be of more use. This is due to the fact that in certain cases it is okay to overlap other elements on a person's body, but the design elements should not be present

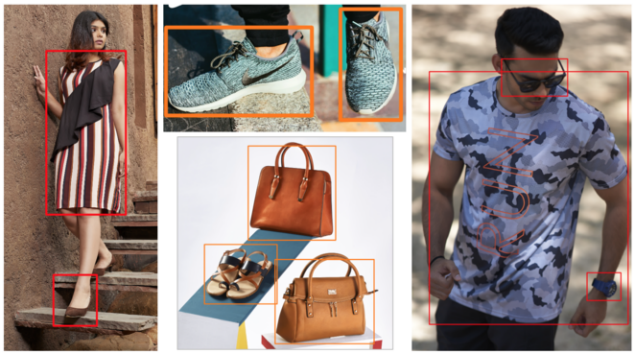

Figure 3: Example images showing how the Fashion Categories detected from the photoshoot images. We can see the detected objects includes regions covering wide range of categories such as footwear, clothing, bags, etc.

on the face. Tagging the gender of the people present will again help filter images based on any particular theme requirement. A CNN-based gender detection model [19] trained on the IMDB-Wiki dataset was used for this purpose.

3.1.4 ***Scene Detection***. Using the scene detector[22], we obtain the various attributes of the background of an image. The categories of scenes include indoor, outdoor as well as details like restaurant, garden, etc and features such as lighting level and man-made area. This level of annotation will help filtering images for both theme based creatives generation and for better personalization.

3.1.5 ***Text Detection:*** We perform text detection on photoshoot images so as to remove few images which have too much text area in the image and are not suitable for generation of creatives. Text was detected using the OpenCV[2] East Text detector.

3.2 Layout Generation

A layout is defined as the set of positions/placements for each of the design elements like "brand logo", "text callouts" etc., on a given content image consisting of people and/or objects along with their bounding boxes. A layout L can be defined as $\{\theta_1, \theta_1 \dots \theta_n\}$ where θ_i represents the co-ordinates of the bounding box for the i^{th} design element. Our objective is to find the co-ordinates θ_i which form the layout with highest aesthetic value.

Layout generation involves evaluating aesthetic compatibility for all combinations of possible locations to properly place and scale the design elements in the banner. Since there are very large number of combinations of the possible coordinates, it is time-consuming to compute to a feasible solution out of all the combinations. For this purpose, we have used Genetic algorithms, as they are suitable for problems involving large combinations of variables. Genetic algorithms have been proven to help in converging quickly to a workable solution and the random mutations also take care of generating new unseen layouts.

Genetic algorithm [13] simulates the natural process of evolution and uses the following techniques to reach the best possible/fittest solution. It starts with an initial population and performs the following steps to reach the best possible solution. In our case, each of the population corresponds to one layout configuration of all the design elements.

a) Selection : Chooses the best parents to produce the next generation from a population

b) Crossover: Combines two parents to create new individuals. In our case, this involves swapping coordinates between the various design elements like text boxes/logo.

c) Mutation: Randomly changes genes/points on individuals to create new individuals. This helps in evaluating new coordinate points.

d) Fitness Function: Uses a fitness score to evaluate the population. Individuals having a higher fitness score are chosen as the next parents.

The Distributed Evolutionary Algorithms in Python (DEAP) library [7] was used for the implementation.

The algorithm generates a random population with x and y coordinates for the logo and text bounding boxes. The bounding boxes of the persons and objects in the photoshoot image is considered as fixed. These coordinates are the inputs for the model. The algorithm performs a series of selection, crossovers and mutations on the logo and text coordinates to come up with the best solution based on the fitness function and the constraints provided. The fitness function incorporates the fundamentals of graphic design by aggregating scores for each of the design aspects. Once a specific number of generations are produced, the individual corresponding to the best score is chosen as the output.

The final fitness/energy score for a layout, $E(X, \theta)$, is the weighted sum of individual scores, $E_i(X, \theta)$. One such layout assessment is showcased in [16]. The weights for the individual fitness scores were obtained by regression of these scores on the CTR of historical banners. We've used CTR, as it helps in decoding user preferences, there by helping us in mapping the weights for different design aspects to user preferences.

$$E(X, \theta) = \sum_{i=1}^{n} w_i E_i(X, \theta)$$

X is the background layout which includes the x and y positions of the bounding box for the person/object in the image. It is denoted as an array $[x_{left}, y_{top}, x_{right}, y_{bottom}]$ which corresponds to the top-left and bottom-right coordinates of the bounding box. θ represents the coordinates for each element and hence is an array of 4 elements. w_i represents the weight for the i^{th} fitness term. Key individual fitness scores are explained below. Note that the overall fitness function can be easily modified to incorporate more design rules.

Alignment: Alignment is one of the core designing considerations while creating a banner. We calculate the misalignment metric which penalizes layouts where the elements are misaligned. For cases with left alignment, we have a lower penalization.

Overlap: Overlapping of any two of the elements, significantly reduces the aesthetic value of the banner. We calculate the overlap percentage for all pairs of elements and penalize them in the fitness score.

$$Overlap\% = \frac{Area_{overlap}}{Area_{total}}$$

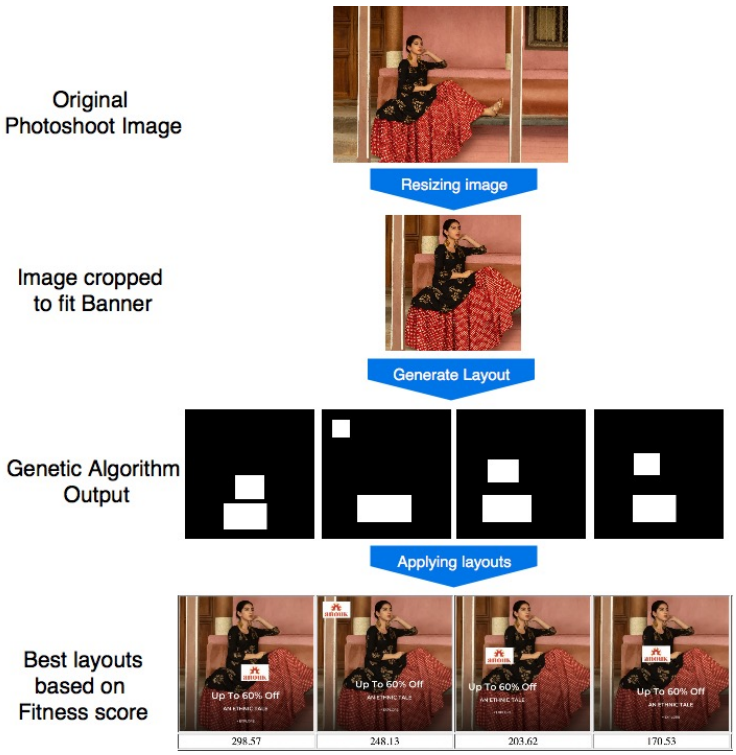

Figure 4: The steps involved in layout generation, and applying them on creatives. The generated layouts and their respective fitness scores are observed. The scores are found to be congruent with the aesthetic value of the banner

Distance between elements: Even in cases of zero overlap between the elements, there can be cases where they are placed very close to each other. This is especially discomforting when the logo or text are placed very close to the face of the person or the important region of an object in the background image. Hence layouts with elements placed farther apart are preferred. The euclidean distance is calculated between pairs and added:

$$Distance(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$
$$\forall (i,j)\ i \neq j$$
$$i,j \in \{Person, Logo, Text, \dots\}$$

Symmetry: Symmetry of the layout is a key factor in stabilizing the overall balance of the final layout. To account for this, we calculate the asymmetry for all elements in the image layout and add this as a penalization term in the fitness score.

$$X_{center} = \frac{(X_{left} + X_{right})}{2}$$

$$Asymmetry_{horizontal} = |2 * X_{center} - Width_{layout}|$$

Constraints: To make sure that all elements are of reasonable size and not too large, the bounding boxes for the elements are assigned a buffer region. All layouts where the dimensions fall outside this region, we term those as in-feasible solutions.

Qualitative Evaluation: The generated layouts are sorted according to their fitness scores and the top one is selected. We carried out an internal survey asking users to identify the designer created layouts and the machine generated ones. It was observed that 60% of the people were not able to distinguish between them. An example result is illustrated in the Figure 4.

For illustration purposes, We have tried to maintain uniformity within single brand/category level creatives and hence the layout that performs best on majority of the cases in further examples.

3.3 Creative Generation

Combining everything together, creative generation involves following steps for a given library of photoshoot images with annotations, brand logos, text callouts.

- Filtering the relevant photoshoot images for the given brand or category using automated tags. Brand name is already provided for the images as labels.
- Automatic cropping of the selected photoshoot images to find the region of interest using annotations.
- Generating best layouts for the given cropped image, logo and text call-outs.
- Overlaying of the different design elements as per the layout specification. Details on how the text is selected and overlaid along with few post-processing steps are explained below.

All the steps involved in creative generation are illustrated for an example in the Figure 5.

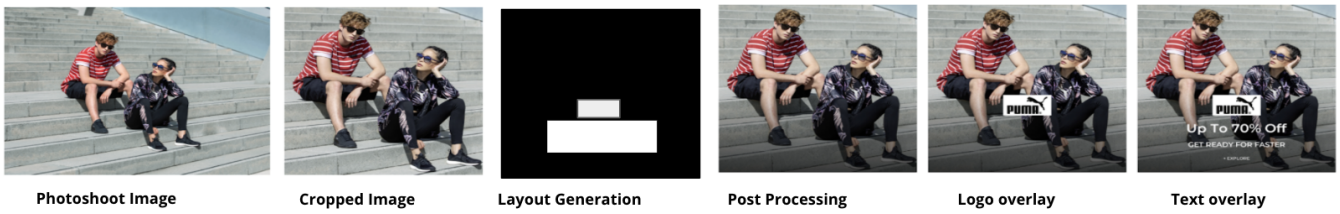

Figure 5: Above figure illustrates the prominent steps involved in creating the banner creative. Given a photoshoot image, it is automatically cropped according to the given requirement of size and category, then best layout is computed using the genetic algorithm based method. After applying image post processing steps, design elements like logo and text call-outs are placed as per the layout specifications.

3.3.1 ***Overlaying text callouts.*** The text on a banner signifies the core message and intention for creating it. Be it a sale or a new arrival, catchy phrases are used to attract the customer. We have a collection of text phrases for various new launches, brand call-outs, sale events. We select the appropriate text to be overlayed from this collection.

Text Formatting. We have a few pre-defined standard text colours, size and fonts based on banner images served on the platform in the past. These were obtained using certain existing design rules, such as golden ratio and the rule-of-thirds. The golden ratio(divine proportion) acts as a typography ratio that relates font size, line height, and line width in an aesthetically pleasing way.

Post Processing. While adding text on an image, we have to ensure that the font color is appropriate. To allow even lighter font colors to be visible, a slight darker gradient is applied on the image.

3.3.2 ***Baseline Approach:*** As a baseline approach for generating creatives, we can crop the center region of the given photoshoot image with required aspect ratio. Further steps include pasting brand logo and text in the same layout as used in the creative generated. Some image processing is done onto the banner creative. Note that this approach doesn't consider regions of any objects/people present in the image. Results using this approach can be seen in Table 3.

3.3.3 ***Qualitative Evaluation of generated Creatives:*** Figure 6 shows examples of banners which were generated by designers and our approach. Figure 10 demonstrates how article-type based creatives are generated. To evaluate the goodness of the generated creatives w.r.t designer created banners, we had conducted an internal survey. In the survey, a user was presented with two images, a designer created banner and automated banner. We found out that only 45% of the people were able to judge correctly, showing that we were able to create banners which are realistic.

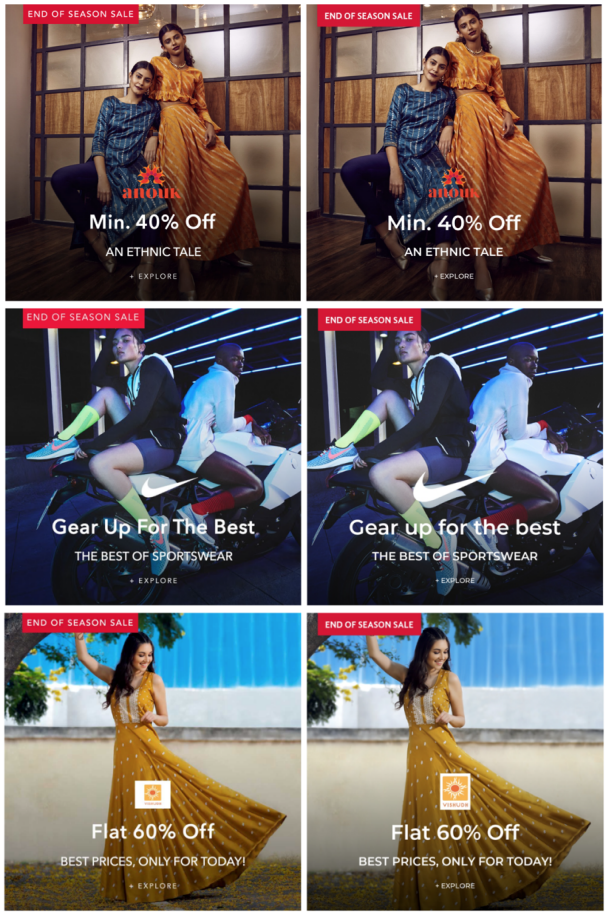

Figure 6: The image on the left was created by a designer while the one on the right was automatically generated.

3.4 Ranking creatives

So far, we have explained ways of automatically generating the banners. Out of these numerous banners generated, we would prefer the superior ones to go live. In order to eliminate the manual effort put in picking these banners, we have designed a ranking methodology based on historical live data-set which had images along with their associated Click-Through-Rate (CTR) as labels. This methodology can be used to rank manual as well as automatically generated creatives. We consider CTR as a signal for goodness of a creative. Note that we do not have any ground truth labels for the generated creatives. We test the goodness of our creatives with the help of the ML model trained on the historical dataset (creatives which are manually designed) which already has a label in terms

of CTR. We have trained different models on this dataset to predict CTR given a generated creative.

Note that the layout generation algorithm explained in earlier sections does the job of finding best layout where as the method explained in this section helps in ranking all creatives and considers features which not only explain about layout, but also the content in the image and its historical performance.

Feature Engineering: The model was built on both image embeddings, explicit features designed from the layout of the image and a set of features determining aesthetic score of the image.

VGG Embeddings. The VGG convolutional network[3] that has been trained on the ImageNet database provides embeddings for an image of dimension 4096. Input to the network requires the image to be re-sized to $224x224$.

Layout Extracted Features. Using the various annotation means used, we can engineer features from the different components of an image.

(1) Position specific features : The coordinates of bounding boxes for people, text, faces and articles.
(2) Area : The relative area percentage covered by each of the dominant objects.
(3) Gender : Number of women present, Number of men present, Total number of people present.
(4) Category Type : The types of articles detected are one hot encoded (present or not) . Types include topwear, bottomwear, watches, etc.
(5) Environment Type : Indoor or outdoor background setting.
(6) Scene Categories and Attributes : Frequently occurring categories (restaurant, garden, etc) and attributes(lighting,man-made, etc.) were picked and one-hot encoded.
(7) Overlapping objects : When text is overlayed on the image, it will overlap on the existing components such as a articles or a person's body. This overlap is tolerable as long as the main article of focus or a face is not hidden. To account for this, the relative area percentage of overlap between each of the following components are calculated :
    - Text regions and Faces
    - Text regions and People
    - Text regions and Articles
(8) Text Quadrants : One hot encoded information for every quadrant if it contains a text component or not.

Aesthetic Features: We have used Neural Image Assessment (NIMA scores)[5] which computes scores representing aesthetics of an image. This is obtained by training a deep CNN on a dataset containing images along with human judgment on aesthetics. In our experiments, we have obtained the score by using this publicly available pre-trained model [5]. For a given image, the aesthetic scores predicted by this model was used as additional feature.

Ranking models: Apart from the simple Logistic Regression model, the tree based classifier were chosen as they are popular methods for CTR prediction problems. Note that other methods based on deep learning [4] could further improve the ranking methodology.

Here are the methods that we have experimented along with the optimal parameters picked.

(1) Logistic Regression
(2) Decision Trees
(3) Random Forest Classifier

As there are fewer clicks compared, we have balanced the data by providing higher weights to samples with clicks.

Ad Personalization Engine: In order to provide a personalized experience, relevant ads are being chosen and shown to the user. For this purpose, all the active ads will be ranked using prediction scores of a click-prediction model. This model is trained on historical click-stream data with user and ad features as input variables. This model is already deployed on production and is not the focus of this work.

4 EXPERIMENTS AND RESULTS

In order to evaluate the goodness of the proposed ranking approach, we have conducted few offline experiments. To evaluate the overall approach, we have performed an online A/B test. Details are explained below.

Dataset. All the images and data used for our experiments are taken from Myntra (http://www.myntra.com), one of the top fashion e-commerce players. We have trained trained the models on nearly 20,000 existing production banner images which already have labels. We perform the feature extraction on all the images in the dataset and then use the target variable as is_clicked. The models were trained on 75% of the data and tested on 25%.

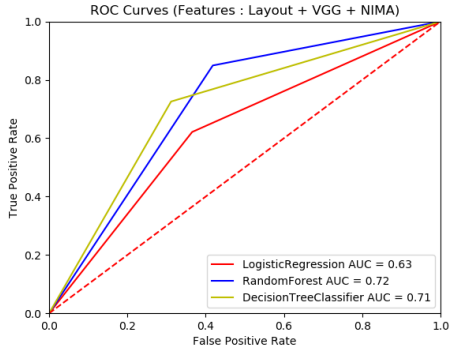

Figure 7: ROC curves for the different classifiers on the combined feature set : VGG embeddings, NIMA and Layout extracted features.

Evaluation Metrics. Classification models were evaluated using both AUC (Area Under Curve) and NDCG (Normalized Discounted Cumulative Gain) [12]. NDCG is a standard information retrieval measure[12] used for evaluating the goodness of ranking a set. For computing NDCG, we have used CTR as relevance score. Evaluation Metrics on test data using different models is presented in the Tables 1 and 2.

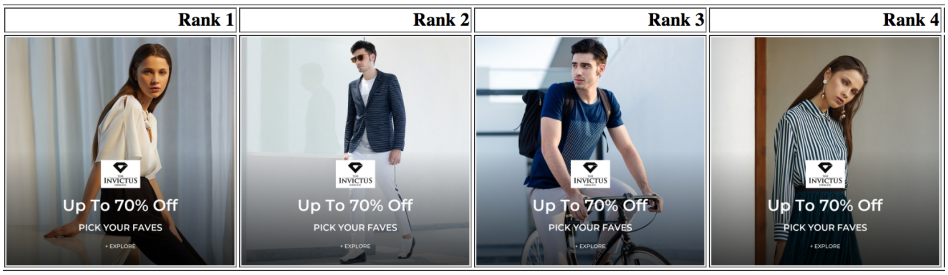

Figure 8: Examples of images with high predicted CTR for the ones generated under a single brand.

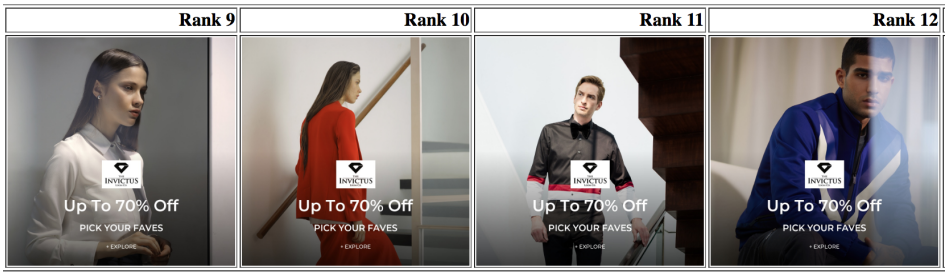

Figure 9: Examples of images with low predicted CTR for the ones generated under the same brand. We found that ranking model is inline with human judgment.

Features	AUC	NDCG
VGG Embeddings	0.71	0.17
Layout extracted features	0.74	0.14
NIMA	0.71	0.24
VGG Embeddings + Layout Features	0.71	0.17
Layout extracted features + NIMA	0.72	0.56
VGG + Layout features + NIMA	0.71	0.22

Table 1: Evaluation metrics for different feature sets experimented on best performing model (Random Forests)

Model	AUC	NDCG
Logistic Regression	0.60	0.03
Decision Trees	0.70	0.05
Random Forests	0.71	0.22

Table 2: Evaluation metrics for different models experimented on combined feature set i.e., VGG embeddings, NIMA and Layout extracted features

Quantitative Evaluation. From the various models and features experimented with, the promising results are shown in Tables 1 and 2. Apart from training on VGG embeddings, NIMA scores[5] and layout extracted features individually, a combination of all was also attempted. Since the best performing model was the Random Forests Classifier, the results for it are present in Table 1. Performance metrics using different models for combined feature set is present in Table 2. We can see that Random Forests using Layout and NIMA features gives the best NDCG. ROC curves for different classifiers using all the feature sets can be see in Figure 7.

It is also interesting to observe that when the model was trained only on the layout extracted features, the most important features were : area of the image, overlap area between text and objects, certain text region coordinates, overlap area between text and people, etc. This further reiterates the fact that the position and orientation of text defines the layout of an image, and it is useful to generate banners without assumptions about their positions.

Online evaluation: We have also performed an online A/B test of the auto-generated creatives along with the manually designed creatives for the same brands. As the main hypothesis was that having multiple options increases personalization, we had a larger set of automated banners in the test set compared to the control set. For both buckets of the A/B test (equal proportion of user traffic), we have used the same personalisation engine that considers both user preferences and banner features. The control group contains manually designed banners and the test group contains automated banners. The results have shown that CTR has increased by 72% for the test set compared to control set (relative increment), with high statistical significance.

Qualitative Evaluation. When the model trained on historical data was used to predict CTR on the newly generated creatives, the results were quite similar to a what the human eye would observe.

Figure 8 is an example of images that have high predicted CTR. This seems to be a meaningful outcome as these images have good color contrast ratio, optimal placement of the components and visible text. In Figure 9, we notice how the images with poor lighting, faces turned away and ones with unnecessary extra padding space

are all pushed down the ranking scale due to much lower predicted CTR.

5 CONCLUSION

In this paper, we have presented a method to generate banner images in an automatic manner on a large scale. This helps in reducing the massive effort and manual hours spent by designers currently and also to produce a wider variety of options to pick from. The broader spectrum of banners will help in catering to wide spectrum of users, instead of showing common banners to all users. We have presented end to end steps involved in generating creatives given few input constraints using automated annotation of photoshoot images. We described a genetic algorithm based method to generate a layout given cropped image and other design elements.

We have shown how a ranking model can be trained on historical banners to rank these generated creatives by predicting their CTR. We observed that when the best performing model was tested on these automatically produced banners, the ranking results was very similar to what a human would've picked, with well positioned, optimal images having higher CTR than the rest. Apart from this offline method of ranking them, future work would be to perform online ranking via reinforcement learning, which will also further boost the personalisation by showing the right set of banners from the vast amount of banners created.

REFERENCES

[1] Fernando Amat, Ashok Chandrashekar, Tony Jebara, and Justin Basilico. 2018. Artwork Personalization at Netflix. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18)*. ACM, New York, NY, USA, 487–488. https://doi.org/10.1145/3240323.3241729

[2] G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[3] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2014. Return of the Devil in the Details: Delving Deep into Convolutional Nets. *CoRR* abs/1405.3531 (2014). arXiv:1405.3531 http://arxiv.org/abs/1405.3531

[4] Junxuan Chen, Baigui Sun, Hao Li, Hongtao Lu, and Xian-Sheng Hua. 2016. Deep ctr prediction in display advertising. In *Proceedings of the 24th ACM international conference on Multimedia*. ACM, 811–820.

[5] Hossein Talebi Esfandarani and Peyman Milanfar. 2017. NIMA: Neural Image Assessment. *CoRR* abs/1709.05424 (2017). arXiv:1709.05424 http://arxiv.org/abs/1709.05424

[6] Michael Fire and Jonathan Schler. 2015. Exploring Online Ad Images Using a Deep Convolutional Neural Network Approach. *arXiv e-prints*, Article arXiv:1509.00568 (Sep 2015), arXiv:1509.00568 pages. arXiv:cs.CV/1509.00568

[7] Félix-Antoine Fortin, François-Michel De Rainville, Marc-André Gardner, Marc Parizeau, and Christian Gagné. 2012. DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research* 13 (jul 2012), 2171–2175.

[8] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587.

[9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask R-CNN. *arXiv e-prints*, Article arXiv:1703.06870 (Mar 2017), arXiv:1703.06870 pages. arXiv:cs.CV/1703.06870

[10] Xian-Sheng Hua. 2018. Challenges and Practices of Large Scale Visual Intelligence in the Real-World. In *Proceedings of the 26th ACM International Conference on Multimedia (MM '18)*. ACM, New York, NY, USA, 364–364. https://doi.org/10.1145/3240508.3267342

[11] Ali Jahanian, Jerry Liu, Qian Lin, Daniel Tretter, Eamonn O'Brien-Strain, Seungyon Claire Lee, Nic Lyons, and Jan Allebach. 2013. Recommendation system for automatic design of magazine covers. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 95–106.

[12] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* 20, 4 (2002), 422–446.

[13] K.S. Tang K.F. Man and S. Kwong. 1996. Genetic Algolithms: Concepts and Applications. *IEEE Transactions on Industrial Electronics* 43, 5 (1996), 519 – 534.

[14] Jianan Li, Tingfa Xu, Jianming Zhang, Aaron Hertzmann, and Jimei Yang. 2019. LayoutGAN: Generating Graphic Layouts with Wireframe Discriminator. In *International Conference on Learning Representations*. https://openreview.net/forum?id=HJxB5sRcFQ

[15] Paridhi Maheshwari, Nitish Bansal, Surya Dwivedi, Rohan Kumar, Pranav Manerikar, and Balaji Vasan Srinivasan. 2019. Exemplar Based Experience Transfer. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*. ACM, New York, NY, USA, 673–680. https://doi.org/10.1145/3301275.3302300

[16] Peter OâĂŹDonovan, Aseem Agarwala, and Aaron Hertzmann. 2014. Learning layouts for single-pagegraphic designs. *IEEE transactions on visualization and computer graphics* 20, 8 (2014), 1200–1213.

[17] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.

[18] Philipp Sandhaus, Mohammad Rabbath, and Susanne Boll. 2011. Employing aesthetic principles for automatic photo book layout. In *International Conference on Multimedia Modeling*. Springer, 84–95.

[19] Yusuke Uchida. 2018. Age-Gender-Estimation. https://github.com/yu4u/age-gender-estimation

[20] Rexroth Xu. 2017. AI visual design is already here. https://medium.com/@rexrothX/

[21] Yunke Zhang, Kangkang Hu, Peiran Ren, Changyuan Yang, Weiwei Xu, and Xian-Sheng Hua. 2017. Layout Style Modeling for Automating Banner Design. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*. ACM, 451–459.

[22] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2014. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*. 487–495.

[23] Ke Zhou, Miriam Redi, Andrew Haines, and Mounia Lalmas. 2016. Predicting pre-click quality for native advertisements. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 299–310.

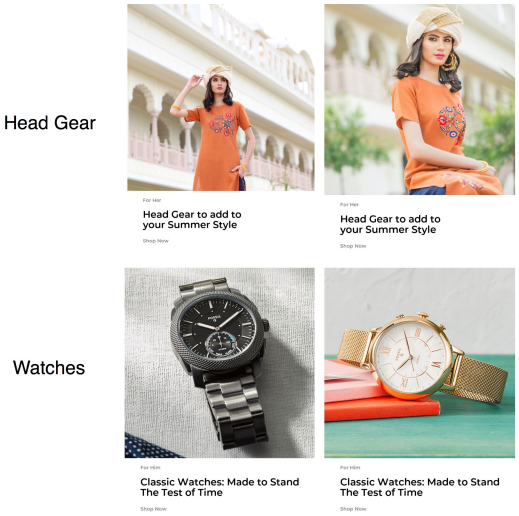

Figure 10: Creatives generated for Article types. First row correspond to the "head gear" category and second row shows creatives generated for "watches" category.

A MORE QUALITATIVE RESULTS

Few examples showing the raw photo-shoot, results using baseline approach and our approach. We can see how faces, body parts can be cut or overlapped with design elements if baseline approach was used.

Table 3: Few Qualitative results with baseline and our approach

Photoshoot Image	Baseline approach	Our approach