

# Exploring Online Ad Images Using a Deep Convolutional Neural Network Approach

Michael Fire  
Sizmek Technologies LTD  
michael.fire@sizmek.com

Jonathan Schler  
Sizmek Technologies LTD  
jonathan.schler@sizmek.com

## Abstract

Online advertising is a huge, rapidly growing advertising market in today's world. One common form of online advertising is using image ads. A decision is made (often in real time) every time a user sees an ad, and the advertiser is eager to determine the best ad to display. Consequently, many algorithms have been developed that calculate the optimal ad to show to the current user at the present time. Typically, these algorithms focus on variations of the ad, optimizing among different properties such as background color, image size, or set of images. However, there is a more fundamental layer. Our study looks at new qualities of ads that can be determined before an ad is shown (rather than online optimization) and defines which ads are most likely to be successful.

We present a set of novel algorithms that utilize deep-learning image processing, machine learning, and graph theory to investigate online advertising and to construct prediction models which can foresee an image ad's success. We evaluated our algorithms on a dataset with over 260,000 ad images, as well as a smaller dataset specifically related to the automotive industry, and we succeeded in constructing regression models for ad image click rate prediction. The obtained results emphasize the great potential of using deep-learning algorithms to effectively and efficiently analyze image ads and to create better and more innovative online ads. Moreover, the algorithms presented in this paper can help predict ad success and can be applied to analyze other large-scale image corpora.

**Keywords.** Machine Learning, Convolutional Neural Network, Deep-Learning, Online Advertising.

## 1 Introduction

Online advertising is one of the largest advertising markets in the world, and it has grown rapidly in recent years [22]. According to ComScore, about 5.3 trillion display ads were delivered in the U.S. throughout 2012 [17]. Furthermore, Magna Global [19] predicts that online advertising will outgrow TV

advertising, which is currently the leading advertising medium in the U.S., and by 2017 online ad revenues will reach 72 billion dollars. Additionally, Forrester forecasts that by 2019 U.S. advertisers will spend over 100 billion dollars on digital advertising, while estimating TV advertising to be only 90 billion dollars [21].

One form of online advertising is web banners (also referred to as banner ads) in which ads are embedded into web pages as static images. The web banners seek to attract traffic to advertisers' websites by prompting website visitors to engage with the ads, mostly by clicking the ads and then being directed to the advertisers' websites. A prevalent method to measure the success of a web banner is measuring the click-through rate (CTR) of an ad by calculating the ratio of the number of times an ad was clicked to the number of times an ad was presented. Moreover, one of the most common advertising revenue models is paying according to ad performance with cost-per-click (CPC) billing [24]. Therefore, predicting an online ad's success becomes fertile ground for research [24, 8, 20].

In this study, we used a large-scale, unique-images dataset, which consisted of 261,752 banner ads from 23 categories, to understand better the world of online advertising. To explore this dataset, we utilized deep-learning algorithms to explore and analyze this dataset in the following manner:

First, we used a trained deep convolutional neural network [7] to identify objects that appeared in each ad. Afterwards, we used the identified objects in each image along with graph theory algorithms to understand the connections among the different ad categories.

By identifying objects that appeared in each ad, we can gain some interesting insights regarding ad categories. For example, we can notice that many ads under the Telecom category contain traffic lights, while many ads under the Gaming category contain space shuttles and pay phones. By recognizing which objects appear under each category, we can better understand the visualization characteristics of image ads in general, and characteristics of image ads in specific categories in particular. Greater understanding promotes improvement and innovation. This type of information can give advertisers recommendations on which objects to embed in their ads in order to make them more appealing, effective, and lucrative.

Second, we used the pretrained deep convolutional neural network to transfer each ad image to its representative vector. Then, we used unsupervised clustering algorithms to divide the ad images into disjointed clusters, which we explored to gain further insights. Using this method, we could determine the main ad banner types that existed in the image corpus.

Lastly, we drilled down to explore web banners that are related to the automotive industry by analyzing an additional image dataset with 34,451 image ads connected with the automotive industry. To inspect this dataset, we transferred each ad into its corresponding vector and divided them into disjoint clusters for exploration. We then utilized the calculated ad vector

representations to create regression models which were able to predict each ad’s CTR.

Throughout this study, we demonstrate the value of deep-learning image processing algorithms in better understanding the domain of image ads. Our methodology provides new insights into this burgeoning field, as well as offering analysis techniques that can be used to reveal significant patterns in other large-scale image corpora. Moreover, these methods can lead to important resources for advertisers wanting to present their products in the most innovative, effective manner.

## 1.1 Contributions

To our knowledge, this study is the first to offer algorithms to analyze a large-scale corpus of image ads by utilizing deep-learning image processing algorithms. Our key contributions presented in this paper are as follows:

- novel techniques to analyze a large-scale categorized image corpus.
- algorithms to infer connections among image categories.
- algorithms for constructing prediction models for ad image CTRs.

## 1.2 Organization

The remainder of the paper is organized as follows: In Section 2, we provide an overview of various related studies. In Section 3, we describe the methods, algorithms, and experiments used throughout this study. Next, in Section 4, we present the results of our study. Then, in Section 5, we discuss the obtained results. Lastly, in Section 6, we present our conclusions from this study and also offer future research directions.

# 2 Background and Related Work

In this study, we primarily utilize three types of algorithms: (a) *deep convolutional neural networks* for processing ad images; (b) *clustering algorithms* for separating the ad images into clusters and understanding the connections between the various ad categories; and (c) *supervised machine-learning algorithms* for predicting an image ad’s CTR. In the rest of this section, we present a brief overview on each one of these types of algorithms.

## 2.1 Deep-Learning Algorithms for Image Recognition

Deep learning is a new area of machine learning in which a set of algorithms attempts to model high-level abstraction in data. One of the common applications of deep-learning algorithms is image processing. By utilizing

deep-learning algorithms to process images, researcher have recognized hand-written digits [3], identified traffic signs [3], detected facial keypoints [28], classified objects in images [15], and more.

For object classification and image categorization there are well-known public datasets, such Caltech-101 [5], CIFAR-10 [14], and MNIST [16], which can be used as benchmarks to evaluate new and existing image processing algorithms. In recent years, deep-learning algorithms have achieved state-of-the-art results on many of these datasets.<sup>1</sup> One of the most popular image processing challenges is the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The ILSVRC Challenge has run annually since 2010 and has attracted participation from more than fifty institutions. One of the two categories of the ILSVRC Challenge is predicting if an object, out of 1,000 predefined object classes, exists or does not exist in an image [25]. In 2012, Krizhevsky et al. [15] used a deep convolutional neural network to classify the 1.3 million high-resolution images in the ILSVRC-2010 training set into the 1,000 different classes. Their trained classifier achieved a top-1 error rate of 39.7% and top-5 of 18.9%. Additionally, Krizhevsky et al. achieved an outstanding top-5 test error rate of 15.3% on the ILSVRC-2012 dataset. In 2014, Szegedy et al., also known as the GoogLeNet team, utilized a convolutional neural network to win first place at the ILSVRC-2014 classification challenge, with a top-5 test error of 6.67% [29]. Recently, He et al. [11] from Microsoft Research achieved a 4.94% top-5 test error rate on the ImageNet 2012 classification dataset. According to He et al., their classifier demonstrated unprecedented results that were “surpassing human-level performance on ImageNet classification.”

There are many deep-learning software frameworks, such Theano [2], Caffe [12], Deeplearning4j,<sup>2</sup> and GraphLab [18], that enable researchers to easily run and evaluate deep-learning algorithms. In this study, we chose to utilize Graph-Lab’s implementation of an image category classifier [9] that was derived from the study of Krizhevsky et al. [15]. Throughout this study, we use the pretrained deep-learning classifier [9] to predict objects that appear in ad images. Moreover, we used the classifier to transfer images to their vector representations.

It is worth mentioning that although deep-learning algorithms, such as deep convolutional neural networks, are very useful and present state-of-the-art results in many image categorization and object classification challenges, these types of algorithms have flaws that need to be kept in mind [30]. For example, Artem Khurshudov recently demonstrated that many implementations of deep convolutional neural network classifiers classified a leopard print sofa image as Felinae images [13].

---

<sup>1</sup>[http://rodrigob.github.io/are\\_we\\_there\\_yet/build/classification\\_datasets\\_results.html](http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html)

<sup>2</sup><http://deeplearning4j.org/>

## 2.2 Clustering Algorithms

During this study, we used two types of clustering algorithms in order to achieve two separate goals. To understand the connections among different ad categories and objects (see Section 3.1) we utilized a community detection algorithm. This type of algorithm organizes graph vertices into communities. Usually, many links exist among vertices in the same community, while comparatively fewer links exist among vertices in different communities. There are many various community detection algorithms [6]. In this study we chose to use the GLay [27] clustering algorithm that is implemented as part of Cytoscape’s clusterMaker2 application.<sup>3</sup>

Our second type of clustering algorithm was used to identify clusters of images according to image properties (see Section 3.1). For this, we used the k-means clustering algorithm [1]. K-means is a widely used clustering algorithm which separates  $n$  observations into  $k$  clusters by seeking to minimize the average squared distance between points in the same cluster. To use the k-means on a dataset one needs to preselect the number of clusters  $k$ . There are various algorithms to identify the recommended number of clusters, such as the gap statistic method that was presented by Tibshirani et al. [32].

In this study, we used the k-means++ algorithm, which augments k-means with a simple, randomized seeding technique that can quite dramatically improve both the speed and the accuracy of k-means clustering [1]. Additionally, we chose the number of clusters  $k$  using a simple heuristic that is described in detail in Section 3.1.

## 2.3 Ad Success Prediction

Displaying the right online ads that will be clicked by a user can greatly influence both the user’s experience and the revenue from ads of advertisers that use a cost-per-click billing model. Therefore, in the last decade, predicting if a user will click an online ad has become fertile ground for researchers.

In 2006, Regelson and Fain [23] introduced a method for improving CTR prediction accuracy using keyword clusters. In 2007, Richardson et al. [24] utilized various features of search ads (i.e., ads that appear mainly in a search engine’s results) to construct a logistic regression model that can predict the click-through rate of new ads. In 2008, Dembczyński et al. [4] utilized the Beyond Search dataset to construct their model for CTR prediction using an ensemble of decision rules. Moreover, Dembczyński et al. demonstrated how their suggested algorithms can be used to provide recommendations in order to improve the ads’ quality. Later in 2011, Wang et al. [33] attempted to predict the ideal number of ads that should be displayed for a given search-engine query.

---

<sup>3</sup><http://apps.cytoscape.org/apps/clusterMaker2>

Recently, the following competitions made public large-scale, CTR-related datasets:

- *2012 KDD Cup*,<sup>4</sup> in which the challenge was to predict the CTR of ads in online social networks given the query and user information.
- *CriteoLabs Display Advertising Challenge*,<sup>5</sup> in which the challenge was to create accurate algorithms for CTR estimation on Criteo’s dataset that contains a portion of Criteo’s traffic over a period of 7 days.
- *Avazu’s Click-Through Rate Prediction Challenge*,<sup>6</sup> in which the challenge was to predict, using Avazu’s released dataset that contains 10 days of click-through data, whether a mobile ad will be clicked or not.

With the release of these datasets, there was an immense amount of interest, over a thousand researchers according to the competition pages on Kaggle,<sup>556</sup> in developing CTR prediction models. In both the CriteoLabs Display Advertising Challenge and Avazu’s Click-Through Rate Prediction Challenge, the winners used field-aware factorization machines (FFM) to achieve the best overall results [35].

### 3 Methods and Experiments

In this study, we utilized two ad image datasets. The first dataset, referred to as *all-ads dataset*, contains 261,752 images<sup>7</sup> collected from advertising campaigns of over 6,500 brands.<sup>8</sup> These images can be divided into 23 unique categories (see Figure 1). The second dataset, referred to as the *auto-ads dataset*, contains 34,451 ad images<sup>9</sup> from over 800 brands that were labeled as ads which are related only to the automotive industry.

#### 3.1 All-Ads Dataset

To better understand the all-ads dataset, we chose to utilize the ImageNet based deep-learning classifier (referred to as the ImageNet classifier) that can classify each image into 1,000 different classes according to the objects that appear in each image [15]. Throughout this study, we chose to use

---

<sup>4</sup><http://www.kddcup2012.org/>

<sup>5</sup><https://www.kaggle.com/c/criteo-display-ad-challenge>

<sup>6</sup><https://www.kaggle.com/c/avazu-ctr-prediction>

<sup>7</sup>Each image in our dataset has its own unique MD5 hash. However, it is worth noticing that the same ad can still appear multiple times with minor changes, such as same images with different resolutions, different file formats, different colors, or even very minor changes in the ads’ texts.

<sup>8</sup>Brand can be a company, or a specific product of a company.

<sup>9</sup>The image ads in the auto-ads dataset were created in a separate process than the ad images in the all-ads dataset. Nevertheless, both datasets share 25,611 ad images with unique MD5 hash.

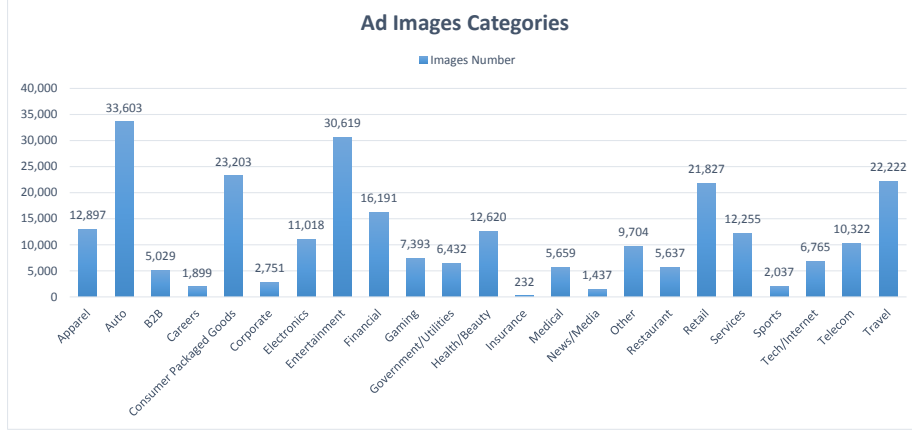


Figure 1: Ad images distributed into 23 categories.

GraphLab’s implementation of the ImageNet classifier [31].<sup>10</sup> For each image in the dataset, we used the classifier to predict the top 5 out of 1,000 object classes that received the highest matching score by the classifier. We then counted the number in each object class (referred to as object) that appeared under each category. During the image analysis, there were cases in which certain objects were repeatedly recognized in many of the ad images. Therefore, similar to the approach of removing stop-words while processing natural language [34], we removed objects that appeared in over 5% of the ads (referred to as stop-objects).

Using the method of counting the number of detected objects in each category can assist in better understanding the types of ads and objects that are used in each specific ad category. However, when using this methodology of counting objects in each category, it is still hard to understand the relationships among the various categories. Therefore, we decided to use graph visualization techniques to view the relationships among objects and categories more clearly. We defined a graph  $G := \langle V, E \rangle$ , where  $V := Categories \cup Objects$  is a set including all 23 ad categories and all the identified objects, and  $E := \{(c, o) | c \in Categories \text{ and } o \in Objects\}$  is a set of links between a category ( $c$ ) and an identified object ( $o$ ), where each object appears in at least 1% of the ads under the linked category. We then used Cytoscape [26] and the resulting constructed graph to visualize the connections among categories and objects. Afterwards, we used Cytoscape’s GLayer [27] community clustering algorithm to separate the graph into disjoint communities, and to reveal connections among the various ad categories.

One goal of this study was to obtain an overview of the various banner ad types which exist in the datasets. To achieve this goal, we used the

<sup>10</sup>GraphLab’s implementation of the ImageNet classifier can be downloaded from [http://s3.amazonaws.com/GraphLab-Datasets/deeplearning/imagenet\\_model\\_iter45](http://s3.amazonaws.com/GraphLab-Datasets/deeplearning/imagenet_model_iter45)

ImageNet classifier to convert each ad image into its corresponding vector in  $\mathbb{R}^{4096}$ . Next, we used the k-means++ clustering algorithm [1] to separate the ad images into clusters. To use k-means++, we needed to predefine the number of clusters  $k$ . To make it possible to manually explore the created clusters, we needed to choose a relatively small  $k$ . However, we still wanted the images in each cluster to be quite similar. Therefore, to identify the most suitable  $k$ , we used the following simple steps: (a) for each  $k \in [2, 50]$ , we used the k-means++ to divide the 261,752 images into  $k$  disjoint clusters;<sup>11</sup> (b) for each  $k \in [2, 50]$ , we calculated the within-cluster sum of squares (WCSS) [10]; (c) to reduce the heuristic’s result variance, we repeated steps *a* and *b* 50 times and calculated the mean WCSS value for each  $k \in [2, 50]$ ; and (d) we selected the  $k' \in [2, 50]$  which presented the lowest mean WCSS value.

We divided the data into  $k'$  clusters using k-means++. Next, we calculated the most common categories, as well as the most common objects in each cluster. We also calculated the average width and the average height of images in the cluster. Afterwards, we randomly selected 50 images from each cluster. Then, we manually reviewed each one of these images to get a sense of the type of ads that belonged to each cluster.

### 3.2 Auto-Ads Dataset

One of the main goals of this study was to validate that deep-learning algorithms, such as a deep convolutional neural network, can be utilized to predict an ad’s success in terms of CTR.

By analyzing a sample of ad images in the all-ads dataset, we discovered that images from different categories tend to have different CTRs. Therefore, to achieve the goal of constructing CTR prediction models, we chose to focus on ad images in a specific category, i.e., the auto-ads dataset. Using the images in the auto-ads dataset, we constructed regression models for predicting image-ad CTRs by initiating the following steps: (a) to use only ad images that had a valid CTR, we removed from the auto-ads dataset images that were used less than 5,000 times, and had a highly exceptional CTR of over 0.2; (b) we used the ImageNet classifier to transfer each ad image into its corresponding vector in  $\mathbb{R}^{4096}$ ; (c) we used Linear Regression, Random-Forest, and Boosted-Tree algorithms which are available in GraphLab’s regression toolkit<sup>12</sup> to construct prediction models that used each image’s corresponding vector, running the set of regression modules with their default values except for the Boosted-Tree algorithm we set the max iterations number to be equal to 100, and for the Random-Forest algorithm we set the number of trees to be equal to 100; and (d) we evaluated the constructed regression

<sup>11</sup>In this study, we used GraphLab’s implementation of the k-means++ algorithm with maximal iteration number set to be equal to 15.

<sup>12</sup><https://dato.com/products/create/docs/graphlab.toolkits.regression.html>



models’ performances by calculating each model’s root-mean-square error (RMSE) value.

One of the factors that can influence ad performance is the ad’s dimensions. In order to better understand this influence, we calculated the Pearson correlations between the ad images’ *width* and the ad images’ CTR, and between the ad images’ *height* and the images’ CTR. Additionally, we calculated the Pearson correlation between the ad images’ *pixel number*, i.e., the multiplication of each image’s width and height, and the ad images’ CTR. Moreover, to better grasp the influence of the ad images’ dimensions and the ad images’ CTR, we repeated the described-above steps, *a* to *d*, twice: First, we constructed regression modules using only the image width and height as features to construct the models. Second, we constructed the regression modules described above using as features each image’s width and height, as well as the image’s corresponding vector (this set of features is referred as the All-Features set).

## 4 Results

In the following subsections, we present the results obtained using the algorithms and methods described in Section 3. The results consist of two parts: First, in Section 4.1, we present the results of analyzing the all-ads dataset according to the methods we described in Section 3.1. Second, in Section 4.2, we present the results of analyzing the auto-ads dataset according to the methods described in Section 3.2.

### 4.1 All-Ads Dataset Results

For each image-ad category, we calculated the most common objects recognized in the ad images within the category. During our analysis, we detected the following stop-objects: (1) book jacket, dust cover (23.82%); (2) score-board (20.95%); (3) packet (18.03%); (4) screwdriver (15.25%); (5) web site (15.05%); (6) digital clock (14.48%); (7) street sign (10.72%); (8) comic book (10.57%); (10) rule, ruler (9.73%); (11) carpenter’s kit (9.7%); (12) band aid (9.61%); (13) ballpoint pen (8.42%); (14) envelope (8.4%); (15) ski (7.95%); (16) menu (6.38%); (17) rubber eraser (6.53 %); and (18) t-shirt (5.66%). These objects were recognized in more than 5% of the ad images’ top-5 objects. Therefore, we removed these objects. Table 1 presents the most common objects that appear in each category after the removal of the stop-objects.

Next, as described in Section 3.1, we created a graph of connections between categories and objects. The constructed graph contains 280 vertices and 1,772 links. However, due to the density of the graph, it is challenging to understand the connections among the categories and the objects. Therefore, we utilized the GLay community detection algorithm to split the cluster into

Table 1: Most Common Identified Objects in Each Ad Category

Categories	Identified Objects	Categories	Identified Objects
Apparel	nail (13.71%)	Insurance	school bus (12.5%)
	face powder (11.29%)		syringe (9.48%)
	wool (10.2%)		slipstick (6.03%)
	analog clock (6.12%)		sweatshirt (5.17%)
Auto	stopwatch (5.14%)	Medical	carton (4.31%)
	sport car (20.17%)		syringe (8.66%)
	racing car (20.07%)		sunscreen (7.9%)
	station waggon (12.76%)		slide rule (6.4%)
B2B	convertible (11.08%)	News/Media	pill bottle (5.74%)
	car wheel (8.82%)		oil filter (5.04%)
	syringe (8.07%)		tobacco shop (9.05%)
	sunscreen (6.9%)		prison (7.1%)
Careers	oil filter (6.32%)	Other	barbershop (5.5%)
	pill bottle (5.49%)		gas pump (4.45%)
	slipstick (4.95%)		organ (4.38%)
	sweatshirt (5.85%)		syringe (5.05%)
Consumer Packaged Goods	sunscreen (5.79%)	Restaurant	oil filter (4.86%)
	ping-pong ball (5.11%)		sunscreen (4.39%)
	pill bottle (5.06%)		gas pump (3.92%)
	syringe (5.06%)		ring-binder (3.92%)
Corporate	sunscreen (10.29%)	Retail	pretzel (8.37%)
	nipple (7.04%)		eggnog (7.97%)
	lotion (6.42%)		hotdog (6.83%)
	oil filter (5.84%)		lipstick (5.94%)
Electronics	pill bottle (5.71%)	Services	paintbrush (5.91%)
	syringe (8.51%)		sunscreen (4.57%)
	lotion (6.65%)		oil filter (4.53%)
	oil filter (6.62%)		tobacco shop (4.47%)
Entertainment	sunscreen (6.62%)	Sports	lipstick (4.0%)
	lipstick (5.13%)		syringe (3.96%)
	lighter (8.58%)		syringe (7.87%)
	oil filter (8.5%)		sunscreen (5.55%)
Financial	binder (7.19%)	Tech/Internet	slide rule (4.87%)
	screen (6.96%)		pill bottle (4.55%)
	syringe (6.47%)		oil filter (3.89%)
	tobacco shop (7.77%)		tobacco shop (7.95%)
Gaming	organ (6.01%)	Telecom	odometer (6.33%)
	barbershop (5.97%)		prison (6.28%)
	prison (5.86%)		organ (5.94%)
	typewriter keyboard (4.8%)		barbershop (5.25%)
Government/Utilities	syringe (8.64%)	Travel	syringe (9.02%)
	sunscreen (6.69%)		ring-binder (6.59%)
	oil filter (5.31%)		slipstick (6.19%)
	pill bottle (5.23%)		crossword (5.51%)
Health/Beauty	ring-binder (4.78%)		tobacco shop (5.35%)
	tobacco shop (9.05%)		syringe (7.79%)
	slot (6.01%)		gas pump (7.55%)
	organ (5.82%)		oil filter (7.14%)
	gas pump (5.48%)		lighter (6.39%)
	odometer (5.17%)		sunscreen (6.14%)
	syringe (7.76%)		sea-coast (6.3%)
	sunscreen (4.87%)		lakeside (4.94%)
	gas pump (4.8%)		sunscreen (4.44%)
	ATM (4.71%)		tobacco shop (4.17%)
	slipstick (4.68%)		syringe (3.84%)
	sunscreen(14.81%)		
	lotion (13.4%)		
	hair spray (9.84%)		
	syringe (7.02%)		
	oil filter (6.85%)		

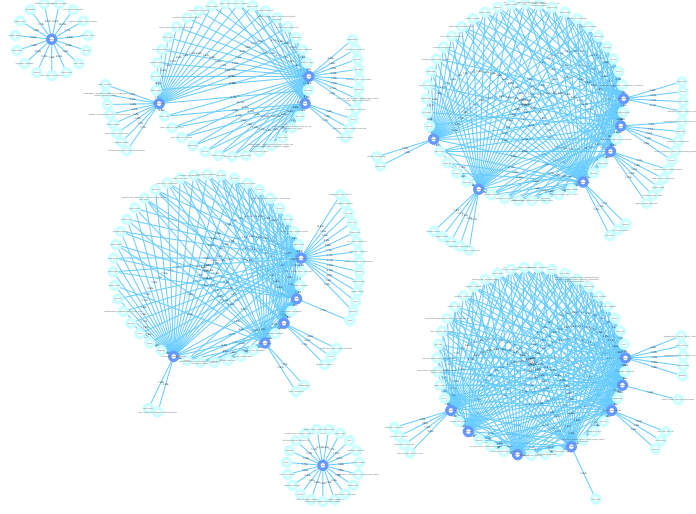


Figure 2: Categories and objects graph. Category vertices are blue, and identified objects are cyan. Each link’s label contains the percentage of image ads in which each object was recognized by the image-processing algorithm in each category. Both the links and the vertices labels are visible by zooming into the graph.

6 communities. Figure 2 presents the six detected communities in which category vertices are marked in blue, while identified objects are marked in cyan. In addition, Figure 3 presents a zoom on the community that included the Auto, Electronics, and Other categories.

Using the ImageNet classifier, we converted each image into its corresponding vector, and we used k-means++ to divide the images into  $k \in [2, 50]$  disjoint clusters according to the method described in Section 3.1. We discovered that using k-means++ with  $k = 14$  presented the minimal mean WCSS (see Figure 4). Therefore, we separated the all-ads dataset into 14 clusters and analyzed each cluster. The results of these analyses are presented in Table 2.

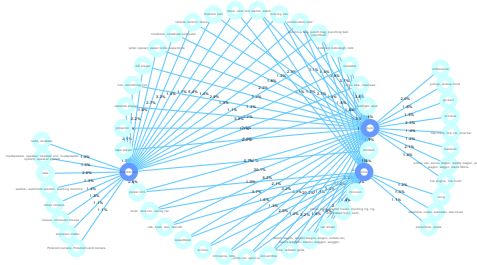


Figure 3: Zoom in on the Auto, Electronics, and Other categories community.

Table 2: Image Clusters using K-means++ (K=14)

ID	Size	Avg. Width	Avg. Height	Top-3 Categories	Top-5 Objects	Description
1	26,891	671.5	83.2	Financial (10.25%)	syringe (35.9%)	Mainly horizontal banner ads. Most of the ads contain the products' brand logo and a short slogan.
				Consumer Packaged Goods (8.95%)	slide rule (15.3%)	
				Retail (8.58%)	screw (15.3%)	
				Services (8.01%)	nail (9.9%)	
				Travel (7.32%)	lipstick (9.5%)	
2	25,727	705	103.2	Auto (19.1%)	lipstick (14.2%)	Horizontal banner ads. In most cases, each ad contains the brand logo and a noticeable image of the product, as well as a slogan.
				Retail (9.71%)	lighter (13.6%)	
				Entertainment (9.67%)	gas pump (12.6%)	
				Consumer Packaged Goods (9.48%)	binder (9.5%)	
				Travel (6.27%)	hair spray (8.9%)	
3	24,577	581.6	375	Travel (22.25%)	nail (9.0%)	Includes ad images with large blank areas. This cluster also includes two side ads, in which the website content appears in the middle, while the ads appear on both sides of the content.
				Apparel (13.0%)	wool (7.2%)	
				Auto (8.19%)	face powder (7.2%)	
				Retail (7.36%)	seacoast (6.7%)	
				Consumer Packaged Goods (7.33%)	lakeside (6.1%)	
4	24,024	826.8	94.3	Entertainment (23.48%)	prison (24.3%)	Mainly horizontal banner ads with dark shaded backgrounds. Most of the ads contain the products' brand logo and people's faces, or cars.
				Auto (12.87%)	organ (23.9%)	
				Consumer Packaged Goods (7.31%)	bell (14.7%)	
				Retail (6.43%)	library (12.0%)	
				Travel (6.24%)	paintbrush (11.6%)	
5	23,337	182.7	596.9	Auto (14.82%)	safety pin (12.9%)	Mostly vertical ads. Many of the ads consist of an image of the product in the center of the ad, and the brand's logo in the top or bottom of the ad.
				Travel (11.27%)	sunscreen (11.9%)	
				Consumer Packaged Goods (10.75%)	hair slide (10.9%)	
				Retail (10.36%)	pill bottle (10.1%)	
				Entertainment (8.39%)	harmonica (9.4%)	
6	19,238	453.2	333.9	Electronics (14.66%)	oil filter (11.2%)	Mostly square ads with blue or white backgrounds. In many of the ads the product is emphasized.
				Auto (11.74%)	sunscreen (10.0%)	
				Retail (9.37%)	lotion (10.0%)	
				Consumer Packaged Goods (9.22%)	whistle (7.6%)	
				Health/Beauty (7.85%)	switch (7.6%)	
7	18,844	302	352.9	Entertainment (27.03%)	ping-pong ball (10.9%)	Various ads which contain people in various poses. There are ads with a single person, a couple, or a group of people. Many of the ads contain the brand's logo and a short text.
				Travel (8.77%)	torch (9.7%)	
				Retail (8.27%)	ballplayer (9.5%)	
				Services (6.3%)	racket (8.5%)	
				Consumer Packaged Goods (5.99%)	tobacco shop (7.9%)	
8	18,232	223.7	542.9	Auto (20.75%)	typewriter keyboard (19.4%)	Mostly vertical banners with dark shaded backgrounds. Many of the ads contain the brands' logos.
				Entertainment (18.74%)	space bar (15.3%)	
				Travel (8.56%)	CD player (8.7%)	
				Apparel (5.55%)	racing car (7.8%)	
				Gaming (5.09%)	tobacco shop (7.6%)	
9	15,963	338.3	301.4	Consumer Packaged Goods (26.07%)	sunscreen (14.6%)	Mainly square banner ads of consumer goods with the product as the focus of the ad. Most ads contain a short text, as well as the brand's logo.
				Retail (13.8%)	oil filter (12.0%)	
				Restaurant (9.36%)	nipple (10.7%)	
				Health/Beauty (6.31%)	plectron (10.7%)	
				Entertainment (6.29%)	lotion (10.0%)	
10	15,298	370.8	372.7	Entertainment (18.39%)	neck brace (17.4%)	Various ads that contain people in various activities, such as playing, walking, and dancing. Most ads contain the brands' logos and a short slogan.
				Apparel (14.07%)	maillot (12.1%)	
				Retail (11.68%)	sweatshirt (11.3%)	
				Health/Beauty (9.67%)	maillot (11.3%)	
				Consumer Packaged Goods (6.14%)	windsor tie (9.2%)	
11	14,819	329	293.9	Financial (13.38%)	sunscreen (23.0%)	Mostly square ads with light backgrounds. Many of them contain the brands' logos, as well as a short text.
				Travel (12.81%)	pill bottle (10.4%)	
				Consumer Packaged Goods (10.06%)	oil filter (9.7%)	
				Retail (8.39%)	carton (9.6%)	
				Services (6.66%)	plastic bag (9.4%)	
12	12,705	449.6	285.3	Financial (12.11%)	hodometer (22.4%)	Horizontal banner ads. In most cases, each ad contains the brand logo and an image of the product, as well as a slogan.
				Electronics (7.81%)	ATM (19.0%)	
				Consumer Packaged Goods (7.28%)	gas pump (12.8%)	
				Travel (6.78%)	oil filter (11.0%)	
				Health/Beauty (6.49%)	screen (8.2%)	
13	12,393	307.7	293.2	Entertainment (23.59%)	tobacco shop (25.6%)	Mostly square ads with dark shaded backgrounds. Many of the ads contain short slogans and numbers.
				Retail (8.05%)	hodometer (18.7%)	
				Consumer Packaged Goods (7.32%)	typewriter keyboard (10.9%)	
				Auto (6.59%)	barbershop (9.2%)	
				Travel (6.11%)	doormat (5.9%)	
14	9,704	379.5	299.2	Auto (88.83%)	sports car (71.2%)	Primarily square car ads in which the car is the focus of the ad. In most ads, the dominant colors of the cars are red, white, black, or blue.
				Other (4.75%)	racing car (62.3%)	
				Retail (1.27%)	wagon (46.1%)	
				Travel (0.55%)	convertible (40.3%)	
				Financial (0.55%)	car wheel (32.1%)	

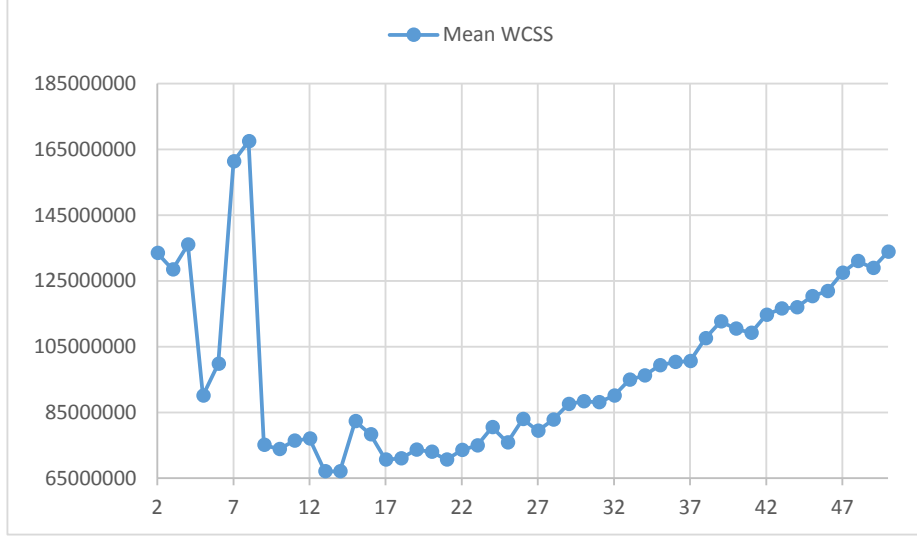


Figure 4: Mean WCSS values for various  $k \in [2, 50]$  values. It can be observed that  $k = 14$  presents the minimal mean WCSS value.

## 4.2 Auto-Ads Dataset Results

As a result of filtering out ad images that were presented fewer than 5,000 times and had a remarkably high CTR of over 0.2, we were left with 12,341 ad images. Next, we calculated the following Pearson correlations between the ad images' CTR and: (a) the ad images' *width* ( $r = 0.1$ ); (b) the ad images' *height* ( $r = 0.147$ ); and (c) the ad images' *pixels number* ( $r = 0.237$ ).

Afterwards, as described in Section 3.2, using the 12,341 ad images, we constructed three regression models using three regression algorithms, and three set of features. Table 3 presents the RMSE of the constructed regression models.

Table 3: Click-Through-Rate Prediction Results

Features	Algorithm	RMSE
Extracted Features	Linear Regression	0.006
	Boosted Tree	0.002
	Random Forest	0.007
Width & Height	Linear Regression	0.0071
	Boosted Tree	0.0049
	Random Forest	0.0071
All-Features	Linear Regression	0.006
	Boosted Tree	0.0019
	Random Forest	0.0072

## 5 Discussion

The algorithms and methods presented throughout this study, which were evaluated on the all-ads dataset and on the auto-ads dataset revealed interesting insights that we will explore in this section.

First, by recognizing the most common objects that appeared in each ad category and are presented in Table 1, we can get a general sense of what types of ads appear in each category. For example, we observe that 20.17% of ads under the Auto category present sports cars, while many ads under the Health/Beauty category present hair spray or various lotions. In addition, we can notice that the technique of removing stop-objects made Table 1 more readable. However, several objects, such as "syringe" and "sunscreen," still appeared as common objects in many of the categories. By manually inspecting over 1,000 randomly selected ads, we did not observe that these objects actually commonly appeared in the ads. Therefore, we believe that the inclusion of these objects is due to false-positive detection of the objects in many of the images. Consequently, to improve the results presented in Table 1, we need to develop better stop-object detection mechanisms. We hope to explore this research direction in a future study.

Second, as can be observed in Figures 2 and 3, by visualizing the graph of links among categories and recognized objects, we can easily examine which objects appear in several categories. For example, we can observe that objects specified as CD player and remote control commonly appear in more than one category. Moreover, we can easily observe which objects are more directly related to a specific category. For example, we can notice that steel bridges and piers commonly appear in ads under the Travel category. This visualization technique can help in quickly understanding the main objects that appear in each category. However, this method has two main disadvantages: One disadvantage is that in the presented visualization, some of the links were removed as a result of the community detection algorithm. Moreover, we only present objects that appeared in at least 1% of the ads under the linked category. Therefore, as a result of this process, some links that may be interesting were filtered out from the constructed graph. The other disadvantage is that the ImageNet classifier is trained to detect only 1,000 different predefined objects. Therefore, if there are other types of objects that are common in image ads, they will not appear in the created graph.

Third, by separating the objects and categories graph into communities using community detection algorithms, it is easier to observe the links among various ad categories that contain similar objects. For example, in Figure 2, we can see that the Auto and the Electronics categories share many objects, such as tape player, CD player, and electric switch. Therefore, these two categories are related to each other. On the other hand, the Consumer Packaged Goods, Telecom, Retail, Health/Beauty, and Restaurant categories

form a separate community which mostly contains different types of objects. By understanding the connections among the various categories we can better understand the image corpus. In a future study, we hope to show that the connections among different ad categories can be utilized to design new creative ads that will be inspired by successful ads in different categories.

Fourth, by dividing the ad images into 14 clusters using k-means++, and then analyzing each cluster (see Table 2), we can reveal several interesting insights regarding the ads in our corpus. We can notice that even though the image ads are separated into different categories, many of the ads under different categories share similar characteristics. For example, most of the ads contain the brand logo, and many of the images contain the brand's slogan or a short text. Additionally, while viewing a sample set from each cluster, we also observed several types of ad images. For instance, we observed the following common types of ads: (a) *text ads*, in which most of the ads contain mainly text, rather than images; (b) *models ads*, in which images of people are the primary focus of the ad, and in many cases these people are performing various actions or standing in various poses; and (c) *product ads*, in which an image of the product is the focus of the ad, something especially common in ads under the Auto category. We also observed that ad images with similar width and height proportions, such as horizontal banner ads, present similar characteristics. Using our clustering, we additionally succeeded in identifying less common types of ads that appeared as part of cluster number 3. Cluster 3 contains ad images that wrap the website content from both sides.

Lastly, according to the results presented in Section 4.2, we found negligible positive correlation ( $r = 0.1$ ) between the images' width and the images' CTR, as well as negligible positive correlation ( $r = 0.147$ ) between images' height and the images' CTR. Additionally, we revealed weak positive correlation ( $r = 0.237$ ) between the images' pixel number and the images' CTR. As expected, these results indicate that larger image ads indeed tend to have higher CTRs. However, as can be inferred from Table 3, we can construct a more accurate CTR prediction model, with an RMSE as low as 0.0019, by using both the size features of the ad as well as the 4,096 features that were extracted using the ImageNet classifier. These results indicate that deep-learning image processing algorithms can assist in predicting an image ad's success.

## 6 Conclusions and Future Work

In this study, we utilized deep-learning image processing algorithms as well as clustering algorithms to explore a large-scale categorized image corpus. We demonstrate that even though the ad-image corpus contains over 250,000 images, our algorithms make it possible to better understand the various types

and layouts of ads that exist in this corpus by sampling only a small subset of corpus, which contains about 1,000 ad images (see Table 2). The methods that are presented throughout this study can also be employed to investigate other categorized and even uncategorized large-scale image corpora to reveal significant patterns and insights. Moreover, by utilizing deep-learning algorithms and extracting the common objects that appear in each category, we show that it is possible to get a sense of which types of objects appear in each category (see Table 1 and Figure 2), and to even understand the connections among the various categories (see Figures 2 and 3). We believe that a better understanding of the connections among the various ad categories and various objects can influence advertisers. They can gain a fresh perspective on the impact of their ads and can discern which elements need to be altered, removed, or incorporated to produce more original and effective ads. For example, ad designers who are considering embedding a specific object into their ad can learn from other ads that have already embedded this object.

According to Section 4.2 results, we can observe that regression models, which utilize features that were extracted using the ImageNet classifier, can predict the CTR of image ads with an RMSE as low as 0.0019 (see Table 3). These models are considerably better than naive models that use only the images’ dimensions to predict the ads’ CTR. This type of prediction model can be instrumental in helping advertisers create more successful ads, resulting in higher traffic to their websites and subsequent increases in sales. Moreover, it can also help them to quickly pinpoint unsuccessful ads and make changes accordingly.

This study is the first of its kind and offers many research directions to pursue. One interesting direction is to improve the image corpus exploration methods presented throughout this study and to apply them on other types of image corpora. Another possible future research direction is to develop algorithms for exploring and analyzing video ads, and for predicting video ad success. A further interesting research direction is to improve the regression models presented in this study by constructing these models with additional features, such as features that include information on the web pages in which the ads were published. One additional possible future research direction includes developing deep-learning algorithms specifically for optimizing the design of image ads. These algorithms can identify objects that, by adding them to an ad, will increase the ad’s performance. For example, these types of algorithms can reveal if adding a black sports car to an ad is better than adding a red SUV. They could even recommend exactly which objects to embed in an image ad in order to directly increase the ad’s CTR. We hope to explore this research direction in a future study.

Overall, the results presented in this study, as well as the offered future research directions, emphasize the vast potential that exists in utilizing deep-learning algorithms in the domain of online advertising. This market is



growing swiftly, and the application of increasingly sophisticated analysis and predictive methods is vital.

## 7 Acknowledgments

We would like to thank Yuval Niv for providing us with the ad image corpus for this article. Additionally, we want to thank Carol Teegarden for editing and proofreading this article.

## References

- [1] D. Arthur and S. Vassilvitskii. k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1027–1035. Society for Industrial and Applied Mathematics, 2007.
- [2] J. Bergstra, F. Bastien, O. Breuleux, P. Lamblin, R. Pascanu, O. Delalleau, G. Desjardins, D. Warde-Farley, I. Goodfellow, A. Bergeron, et al. Theano: Deep learning on gpus with python. In *NIPS 2011, BigLearning Workshop, Granada, Spain*, 2011.
- [3] D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642–3649. IEEE, 2012.
- [4] K. Dembczynski, W. Kotlowski, and D. Weiss. Predicting ads click-through rate with decision rules. In *Workshop on targeting and ranking in online advertising*, volume 2008, 2008.
- [5] L. Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70, 2007.
- [6] S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.
- [7] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.
- [8] T. Graepel, J. Q. Candela, T. Borchert, and R. Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft’s bing search engine. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 13–20, 2010.

- [9] J. Gu. Training a deep convolutional neural network on imagenet. [https://dato.com/learn/gallery/notebooks/build\\_imagenet\\_deeplearning.html](https://dato.com/learn/gallery/notebooks/build_imagenet_deeplearning.html), November 2014. [Online; accessed 06/22/2015].
- [10] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Applied statistics*, pages 100–108, 1979.
- [11] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *arXiv preprint arXiv:1502.01852*, 2015.
- [12] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.
- [13] A. Khurshudov. Suddenly, a leopard print sofa appears. <http://rocknrollnerd.github.io/ml/2015/05/27/leopard-sofa.html>, May 2015. [Online; accessed 06/22/2015].
- [14] A. Krizhevsky and G. Hinton. Convolutional deep belief networks on cifar-10. *Unpublished manuscript*, 2010.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [16] Y. LeCun, C. Cortes, and C. J. Burges. The mnist database of handwritten digits, 1998.
- [17] A. Lipsman, C. Aquino, and S. Adamo. 2013 u.s. digital future in focus. *ComScore*, February 2013. [Online; accessed 06/09/2015].
- [18] Y. Low, J. E. Gonzalez, A. Kyrola, D. Bickson, C. E. Guestrin, and J. Hellerstein. Graphlab: A new framework for parallel machine learning. *arXiv preprint arXiv:1408.2041*, 2014.
- [19] Magna Global Advertising Research. Magna global’s us ad forecast update. [http://news.magnaglobal.com/article\\_display.cfm?article\\_id=1652](http://news.magnaglobal.com/article_display.cfm?article_id=1652), August 2014. [Online; accessed 06/09/2015].
- [20] H. B. McMahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, et al. Ad click prediction: a view from the trenches. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1222–1230. ACM, 2013.

- [21] L. O'Reilly. The era of tv's media dominance will come to an end in 2016 — here's the evidence. <http://www.businessinsider.com/digital-to-overtake-tv-advertising-in-2016-2014-11>, November 2014. [Online; accessed 06/09/2015].
- [22] PwC. Iab internet advertising revenue report 2014 full year results. [http://www.iab.net/media/file/IAB\\_Internet\\_Advertising\\_Revenue\\_FY\\_2014.pdf](http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_FY_2014.pdf), April 2015. [Online; accessed 06/09/2015].
- [23] M. Regelson and D. Fain. Predicting click-through rate using keyword clusters. In *Proceedings of the Second Workshop on Sponsored Search Auctions*, volume 9623, 2006.
- [24] M. Richardson, E. Dominowska, and R. Ragno. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*, pages 521–530. ACM, 2007.
- [25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, pages 1–42, 2014.
- [26] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.
- [27] G. Su, A. Kuchinsky, J. H. Morris, F. Meng, et al. Glay: community structure analysis of biological networks. *Bioinformatics*, 26(24):3135–3137, 2010.
- [28] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3476–3483. IEEE, 2013.
- [29] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*, 2014.
- [30] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.
- [31] P. Teterwak. Deep learning: Doubly easy and doubly powerful with graphlab create. [http://blog.dato.com/deep-learning-blog-post?\\_ga=1.123380683.1189688547.1424700181](http://blog.dato.com/deep-learning-blog-post?_ga=1.123380683.1189688547.1424700181), November 2014. [Online; accessed 07/20/2015].

- [32] R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423, 2001.
- [33] B. Wang, Z. Li, J. Tang, K. Zhang, S. Chen, and L. Ru. Learning to advertise: How many ads are enough? In J. Huang, L. Cao, and J. Srivastava, editors, *Advances in Knowledge Discovery and Data Mining*, volume 6635 of *Lecture Notes in Computer Science*, pages 506–518. Springer Berlin Heidelberg, 2011.
- [34] W. J. Wilbur and K. Sirotkin. The automatic identification of stop words. *Journal of information science*, 18(1):45–55, 1992.
- [35] J. Yu-Chin, C. Wei-Sheng, and Z. Yong. Libffm: A library for field-aware factorization machines. <http://www.csie.ntu.edu.tw/~cjlin/libffm/>, April 2015. [Online; accessed 06/22/2015].