Exploring the links between creative execution and marketing effectiveness.

Ekimetrics. × Meta

Contents

Measuring the efficiency of creativity in advertising has historically been challenging.

It is hard to isolate the impact of creativity from other factors that impact performance, such as execution tactics or brand health. As the advertising landscape evolves, with brands using several creatives at once and with more and better data, it has become increasingly important to understand the impact of creativity.

This paper provides a technical Marketing Mix Modelling (MMM) approach (object detection algorithms and multi-stage econometric modelling) that demonstrates an objective approach to creative measurement.

We used a sample of five brands from the Insurance, Cosmetics, Hospitality and automotive sectors, and 13 related outcome KPIs. We concluded that there are commonalities around what features make a particular creative drive a higher return on investment, and what features are irrelevant. People and Product in isolation and combined, are the features that when appearing on Meta creatives, drive the highest ROIs.

> **“People and Product in isolation and combined are the features that drive the highest ROIs.”**

Summary of Research Statistics

Brand	Total
KPI’s	4
Brands and Sub-Brands	10
Models	124
Timeframe	3.5 years
Campaigns	2.3K
Creatives	22.8K
Spend (USD)	8.9M
Impressions	3.0B

Methodology-Related Learnings:

1

Machine Learning (Object Detection) technical MMM is an objective approach to creative measurement, beyond execution tactics or brand power.

2

A combination of pre-trained and custom-trained models to detect objects within creative assets are needed for this type of analysis. These require human resources and cloud computing power at large scale.

3

Custom-trained models are required to detect brand-specific objects, such as logos, brand cues, and products.

4

Pre-trained models are useful for detecting generic objects such as cars or food, but the default models may not always be optimal, and further optimization may be necessary to improve detection in custom datasets.

5

The distribution of labels across creatives is important to be able to measure efficiency – brands with a very high % of the same features proved difficult to obtain robust results.

6

There are outstanding limitations to what Object Detection can do, with some brands being more affected in the labelling of features specific objects. For example, the detection of the interior of cars was limited.

Feature-Related Learnings:

7

The most efficient features at driving higher ROIs across all observed brands were Product and Person with a Face. When combined, Person & Product were also shown also to be important.

8

Person & Smiles were more efficient in brands selling intangible products, rather than physical products.

9

Text was irrelevant in most brands. Except for Cosmetics, where combining a product with text was shown to drive an ROI uplift vs. product on its own which had no significant uplift.

10

Logo & Brand Cue were efficient at driving higher ROIs for the Hospitality brand. Interestingly, for Insurance, Brand Cue was relevant but Logo was not.

The link between creative execution and marketing effectiveness presents measurement challenges and is therefore poorly understood

- There are no rules on how creatives should be distinguished, and subjectivity also means one person's definition of 'good creative' will be different from the next. Third parties exist for testing creative, but the link between their metrics and sales performance is not strongly established
- There is an accelerating trend towards brands using more creatives, often tailored to specific audiences, vs fewer creatives in traditional advertising. This means it is increasingly difficult to identify a common creative thread to capture and understand performance
- The above statement is true for all marketing channels but particularly true for digital channels and Meta platforms
- Collaboration between marketers, creative and analytics teams is essential to delivering strong insights. However, this can be difficult as those teams rarely work together. It means more data, more conversations and building relationships. This paper goes some way to facilitating and overcoming those barriers.
- MMM is one of the most common forms of marketing measurement and the technique has not evolved to capture creative well
- Brands could optimise their marketing returns further if they could uncover insights about which creatives perform well and resonate with their audience
- An objective MMM approach to creative measurement is crucial to fit this new creative landscape.

The methodology from this study may be applicable for many different audiences:

- Marketers can answer questions such as "how can marketing returns be improved through creative execution?"
- Marketing effectiveness professionals can answer questions such as "how can we codify the creative elements of marketing campaigns?", "how can I improve measurement when varied creatives are used simultaneously?"

The methodology used in this study can be applied as a whole, or to answer questions relating to Object Detection (OD), like:

- Sponsorship content detection: OD Methodology can be applied to video footage of sports games or other material to detect sponsorship content
- Product Placement: OD Methodology can be applied in combination with scraping methods to detect organic product or logo placements. This could be in social media or other scrapable media sources
- MMM Sub-Modelling of Meta or other activity: MMM Methodology can be adapted for other investigations into Meta activity, such as creative type, creative placement, format, caption/messaging in creative, etc. The same methodology can also be applied to other social media platforms like YouTube or TikTok.

The creative elements of a campaign are defined by features. Features are objects or sets of objects present in the marketing creative, all of which are identified by object detection algorithms.

This includes common objects like people, as well as brand-specific objects, such as a logo or product.

Each creative is then codified according to the set of objects, allowing for the creation of time series variables that can be modelled through MMM.

Using 14 MMM Models as a base of this analysis, alongside creatives and their impressions, iterations of sub-models were generated. These test different features or combinations of features with the objective of understanding:

- What features are relevant to a creative?
- What features (or combinations of features) drive the higher Returns on Investment?

The methodology follows a 5-step approach:

1. Data
2. Feature and Object Selection
3. Object Detection
4. Feature Engineering
5. MMM

Sufficient and relevant data must be gathered for each brand, whereby the impressions in the MMM models match the impressions in the creative data.

If your coverage is above 100%, you probably have duplicated creatives. The Convolutional Neural Network (CNN) method from the Python library imagededup is efficient at de-duplicating data.

Different creative types can be used; in this analysis we used Static, Carousel and Video creatives. Video creatives must be pre-processed. Every 10th frame from videos was extracted, resulting in between 30,000 and 90,000 image files for each brand.

All the Data pre-processing was done in Databricks.

Dataset Name	Type	Source	Contents	Purpose
Creatives	Un- structured	Meta	Raw image, video and carousel creatives posted on Facebook and Instagram by the brands	Input data for the object detection process
Creative Body	Cross-Sectional Data	Meta	Caption text and flag for carousel for each creative	To distinguish static images from carousels; To define the presence of a celebrity in the creative
Spend & Impressions	Panel Data	Meta	Daily spend and impressions associated with each individual creative	Independent variables for the MMM sub-models
Modelling	Panel Data	Ekimetrics Clients	Weekly base, market, seasonality and media variables, along with the KPI of interest	Dataset for the MMM base models

See Appendix: Objection in more detail for further information.

The objects used to encompass the creative elements across the four brands are summarized in Table 1.

The objective of the selection process is that the features are meaningful to customers, that these appear in creatives from different brands and that these are available in the data. Feasibility and frequency of appearance is crucial.

Table 1: Feature Selection

Feature	Definition	Object Detection Model
Logo	Brand-specific logo	Custom-Trained Faster R-CNN
Brand Cue	A visual cue that resembles the brand logo or encompasses the brand DNA	Custom-Trained Faster R-CNN
Person	Any part of a person, or multiple people	Pre-Trained Faster R-CNN
Face	A person with their face shown	Custom-Trained Faster R-CNN
Smile	A person who is smiling	Custom-Trained Faster R-CNN
Product	Brand-specific product	Pre-Trained and Custom-Trained Faster R-CNN
Text	Text in the creatives, excluding text on products	Optical Character Recognition

*R-CNN = region-based Convolutional Neural Network

After determining the broad categories of objects to include in the study, a more involved process was followed for defining the exact objects to be identified.

Table 2: Objects identified by feature and brand

	Insurance	Automotive	Cosmetics	Hospitality
Person	Person	Person	Person	Person
	Smile	Smile	Smile	Smile
	Face	Face	Face	Face
Product	Technology	Car exterior	Perfumes	Food
	-	Car interior	Perfume boxes	Furniture
Logo	Each brand had their own logo			

These are shown in Table 2 for each brand, and Figure 1 illustrates examples of what features and combinations of features would look like in actual images.

Figure 1. Example Images and Associated Features.

Person & Product (No Face)

Product & Logo

Logo Only

Person & Face & Product /
Person & Face & Smile

Products Only

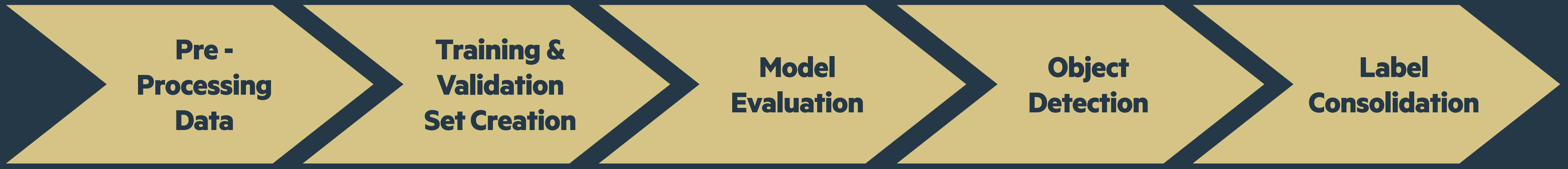
Person & Face Only / Brand Cue Only

Object Detection

The object detection process relied on a combination of pre-trained and custom-trained models from two Python libraries: Tesseract and Detectron2.

Tesseract, an open-source Optical Character Recognition (OCR) library, was used for detecting text, while Detectron2, Meta AI Research's next-generation platform for object detection and segmentation, was used for the remaining objects. From Detectron2, pre-trained object detection models were used for detecting common objects such as people, cars, and technology, while custom-trained models were used for detecting brand-specific objects such as logos. Figure 2 outlines the objects detected by each model type.

The Object Detection workflow is illustrated and explained below:



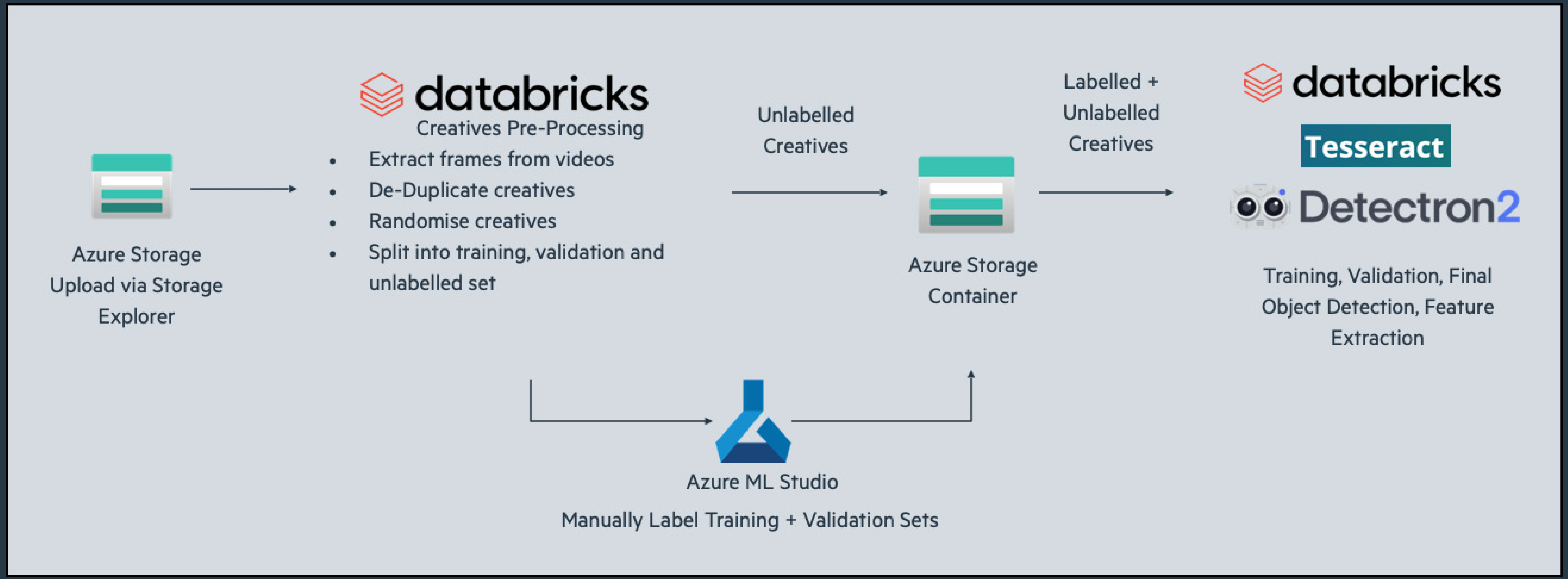
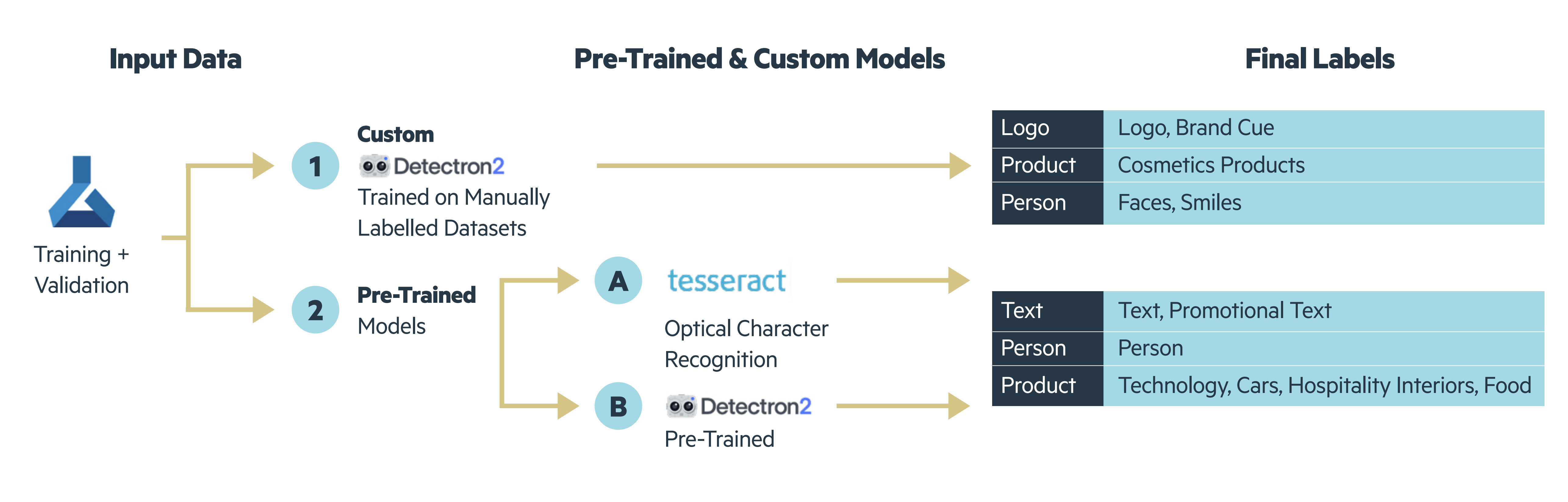


Figure 2. Object Detection using different pre-trained and custom-trained models.



During the feature engineering step, the panel data, which would serve as the independent variables for the sub-models in the marketing mix models, is created.

That is, a dataset containing the total sum of impressions and spend associated with each object per day per brand. While only the impressions data was required for the modelling, the spend was used for deriving the ROIs and thus was also needed at an object level.

Due to the significant overlap in objects appearing in creatives, e.g. many creatives contained both Logo and Product, using this dataset in the sub-models would lead to a poor model performance and would not allow for meaningful insight into the effect of each individual object. Therefore, a separate sub-model was run for each object (feature) or combination of objects.
See MMM section for more detail.

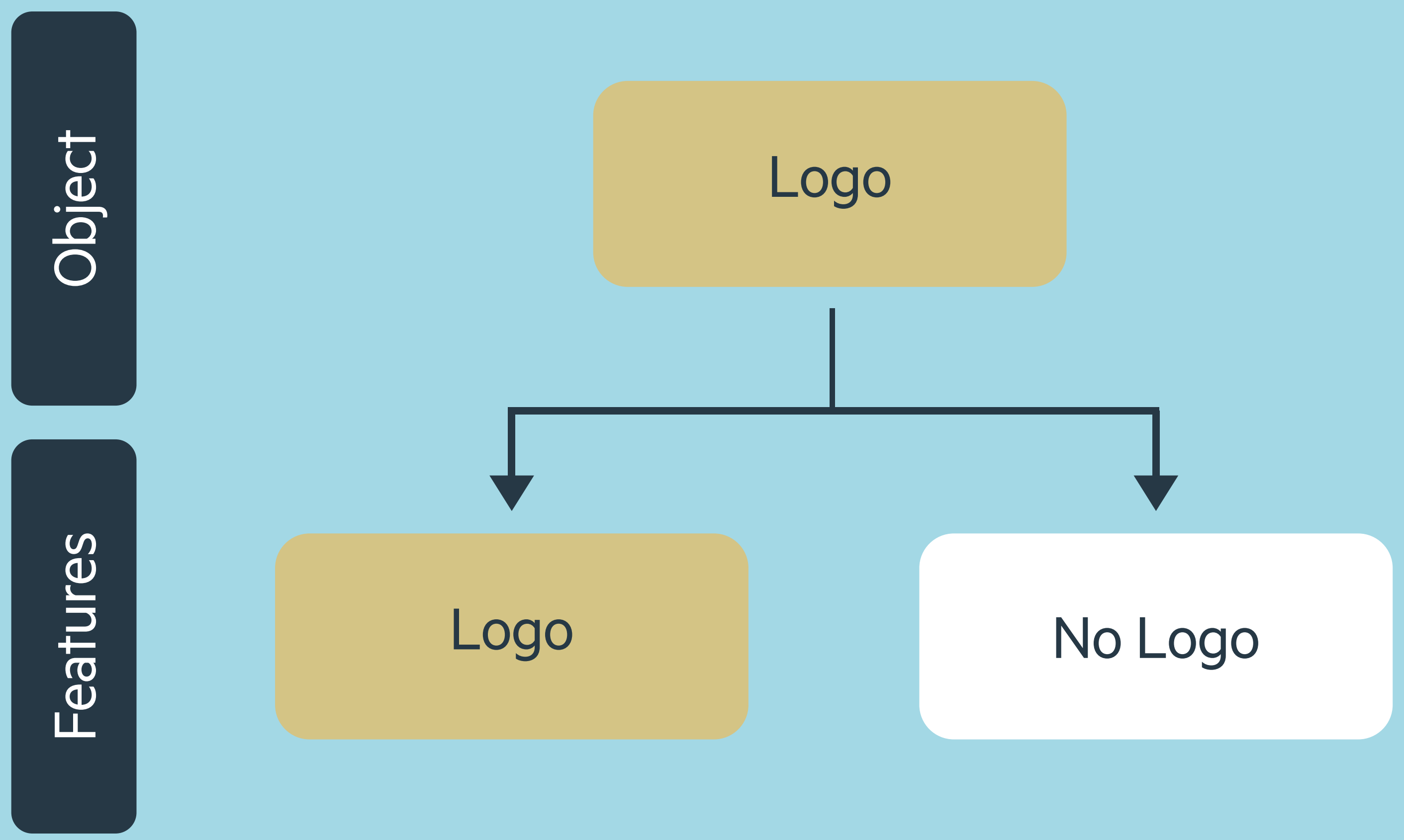
For example, Logo was modelled separately from Product. By doing this, each object had an associated contribution which could be directly compared to that of other objects. In order to do this, each object variable also had to have a partnering variable which captured the impressions **not associated** to that object, so that the effect of including the object could be compared with the effect of not including it. This was done for each of the objects detected (see table 2 for a complete list). Figure 3 illustrates this concept.

Table 3: Distribution of impressions for feature by brand.

Submodel	**Hospitality**	**Insurance**	**Cosmetics**	**Automotive**
Brand Cue	54%	37%	3%	0%
Logo	36%	30%	64%	74%
Text	42%	83%	17%	62%
Logo Size (S/M/L)	44/8/2%	29/2/0%	57/7/1%	67/6/1%
Logo & Brand Cue	9%	27%	1%	0%
Product	43%	6%	67%	83%
Product & Logo	2%	4%	55%	63%
Person	64%	83%	26%	56%
Person & Face	49%	76%	13%	25%
Person & Face & Smile	24%	61%	9%	14%
Person & Product	35%	6%	9%	29%
Person & Face & Product	17%	6%	7%	5%

For example, for the Hospitality brand, 54% of the Meta impressions across all the creatives feature a Brand Cue. The percentages sum to more than 100% because an impression can be linked to creatives containing more than one feature.

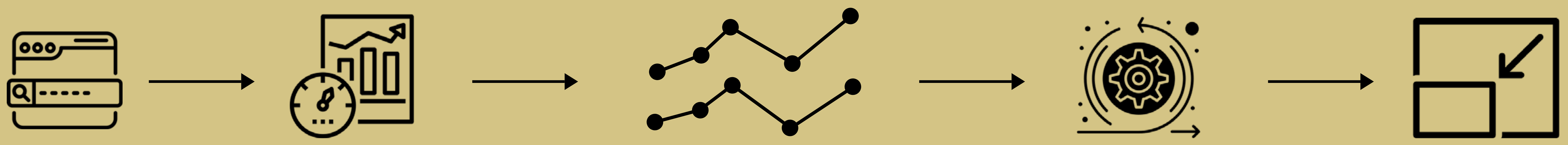
Figure 3: Creation of Logo Features and Partnering Features



An optimized two-staged modelling approach was applied to existing MMM models developed by Ekimetrics for a range of brands, products and KPIs.

First, the base models were optimized, and second, sub-models were created to explain the variation in Meta contribution using the creative features. Bayesian optimization was used in both steps for variable transformation in relation to lag, adstock and saturation.

Figure 4: Modelling Workflow



MMM Models

- Sales
- Quotes
- Web Traffic
- Consideration
- Etc

MMM Model Optimisation

- Hyper-parameterization of Meta variables to enhance performance of all models
- Parameters define Ad-stock, Wei-bull and lag transformations

Sub-Modelling Meta KPI Creation

- Calculating Meta Contribution to each of the enhanced MMM models and residuals
- Selection of ‘trusted’ periods

Feature Sub-Modelling

- Sub-modelling for each of the Meta KPIs
- Iterative process to generate each features/ combination of features ROIs

ROI uplift normalization

- Scaling of features ROI by % of impressions in the different models
- Analysis of absolute and uplift figures to identify feature performance

The MMM models used in this analysis are multilinear regressions accounting for all significant factors driving the KPI (most commonly sales), including Social Media.

The selection of these models was based on the criteria of model stability (high R^2, no autocorrelation, controlled heteroskedasticity) and enough variation in social activity (spend and impressions threshold of 5% per object). Some of the modelling periods were cut to account for sparse data, low correlation between creative impressions and modelled impressions or low variation. Therefore, a trusted relationship from social to the KPI of each MMM was established but could be enhanced by choosing more accurate parameterization of Meta activity through MMM model optimization. See table 4 for details of the models included in the study.

Table 4: Detail of Models selected:

Brand	Sub-brand	KPI	No. of Independent Vars in Base Model	Modelling Period	Base Model R^2
Hospitality	A	Sales	1	74 weeks 28.05.2018 - 21.10.19	97%
	B	Sales	2	92 weeks 11.06.2018 - 09.03.2020	96%
Insurance	A	Sales	5	45 weeks 21.09.2020 - 26.07.2021	70%
	A	Traffic	6		88%
	B	Quotes	3		84%
Cosmetics	-	Sales Consideration	1	117 weeks 27.06.2016 - 24.09.2018	96%-98%
Automotive	A	Sales	1	183 weeks 01.01.2018 - 28.06.2021	87%
	B	Sales	1		79%
	C	Sales	1		84%
	D	Sales	1	122 weeks 04.03.2019 - 28.06.2021	85%

Once the MMM models were optimized, we isolated the impact of Meta on the KPI according to the following relationship:

Meta Contribution to KPI = Coefficient of Meta Variable x Transformed Meta Variable

The Meta contribution became the dependent variable of the Sub-Model, along with the constant and residuals. The object detection impressions timeseries served as the explanatory variables. The objective of this was to understand the impact of different feature splits into the overall Meta performance, allowing room for movement through the constant and residuals.

A sub-modelling approach allowed for the object detection features to have indirect effects on the KPI, through the Meta variable. Bayesian Optimization Methods were employed to find the optimal transformations for each variable.

Each Sub-model is a linear regression, testing the impact of the appearance of a feature, or feature group, and its opposite (partner feature) against the total Meta performance. Since the features are not mutually exclusive (between 48%-66% of creatives contained at least two objects), they could not all be tested in the same model.

Furthermore, the daily impressions and spend per feature were split according to creative type (static, carrousel or video). That is, the impressions associated with a feature or combination of features, e.g., Person & Product, were split out by static, video and carousel. While it was not in the scope of the study, doing this allowed for further analysis into the impact of features by creative type.

A sequential approach was followed in which each feature group was tested in isolation. Within each sub-model, the transformations of the features (lag, adstock, and diminishing returns) which maximized R^2, and minimized p-values were chosen. The transformations that the Meta features follow were bounded to the following ranges: lag of 0-3, adstock of 0-50, saturation (K; S) of 0.1-0.8; 1-9.

An example of sub-model is shown below.

$$\textit{Meta Driven Sales} = \beta_0 + \beta_{1,S}x_{1,S} + \beta_{2,S}x_{2,S} + \beta_{1,v}x_{1,v} + \beta_{2,v}x_{2,v} + \beta_{1,c}x_{1,c} + \beta_{2,c}x_{2,c} + \varepsilon$$

Where:

$x_{1,S}$ = *Creative Impressions when a Brand Cue appears in Static Creatives*

$x_{2,S}$ = *Creative Impressions when a Brand Cue does not appear in Static Creatives*

Same for Video (V) and Carousel (C)

One major challenge with the approach of testing the appearance of a feature or a set of features against its partner feature (e.g. impressions of creatives where a Brand Cue appears vs impressions of creatives where a Brand Cue does not appear), is that the partner feature is indirectly testing the appearance of other features or no features at all.

To manage the risk of results inaccuracy we made sure that:

1. Both the main feature and the partner feature in the sub-models had to be statistically significant at a 10% level. For example, if the appearance of a feature is significant but the partner is not, the results of that regression were not used.
2. The percent of Impressions of the feature being tested had to be > 5%

The last step of the analysis centred around calculating ROIs and deriving ROI uplifts for each feature per brand, using the feature contributions from the sub-models and the associated spend.

This involved the following steps:

1. Calculate the ROI of each feature (and partnering feature/s) using the contribution and the spend associated with that feature. Because the features were modelled by creative type (static, carousel, and video), the results were weighted by the percent of impressions each type represented over the total of the feature.

$$ROI_{feature\ A} = \frac{\begin{array}{c}\text{Incremental KPI driven by}\\ \text{feature A static impressions}\end{array} + \begin{array}{c}\text{Incremental KPI driven by}\\ \text{feature A carousel impressions}\end{array} + \begin{array}{c}\text{Incremental KPI driven by}\\ \text{feature A video impressions}\end{array}}{\begin{array}{c}\text{Investment in impressions with creatives}\\ \text{featuring A}\end{array}}$$

2. Calculate the ROI of the partnering variable (e.g. No Product)

3. Calculate the Uplift:

$$ROI\ Uplift = \frac{ROI\ of\ Feature}{ROI\ of\ Partner\ Feature\ for\ Feature\ Group\ impressions\ were\ the\ feature\ isn't\ present=}$$

4. Index and rank ROI uplifts

a. ROI uplifts are standardized, so they are comparable across brands

b. They are then ranked so that #1 is the feature with the greatest uplift (maximum difference between the feature appearing and not appearing).

With this methodology, we were able to answer the following questions for each brand:

1. Which features are important?
2. Which features are the most efficient?

Which features are important?

In this study, a feature is considered relevant if:

- Splitting the impressions by the appearance of a feature improves the performance of the sub-model
- There is a positive uplift between the feature appearing and not appearing
- If the features in the sub-model (both the feature and partner of the feature group) are significant with a p-value < 0.1.

Table 5 outlines the findings in this study regarding features that matter for each sector. As a reminder, within each sector one or more brands were modelled through numerous sub-models as explained earlier.

Table 5: Features that matter for each brand

Submodel (Feature Group)	Hospitality	Insurance	Cosmetics	Automotive
Brand Cue	✔	✔		
Logo	✔		✔	
Text			✔	
Logo Size (S, M, L)	✔		✔	
Logo & Brand Cue	✔	✔		
Product	✔	✔		
Product & Logo				
Person			✔	
Person & Face	✔			✔
Person & Face & Smile		✔		
Person & Product		✔		✔
Person & Face & Product		✔	✔	

* Automotive results were achieved with further sub-modelling where for Person it does not test Person appearing Y/N but Person with a Face and without a Smile appearing vs not; for Person & Product it tests person without a Face & Product vs not.

Brand Cue, Logo, Product and Person were the most common features that significantly drove ROI uplifts across brands.

Brand Cue was significant at driving higher ROI in Hospitality and Insurance. It could, however, not be tested for Cosmetics and Automotive because of low impression levels (3% and 0% respectively).

Logo was significant in Hospitality and Cosmetics. For Automotive, 74% of creative impressions featured a Logo (either on the car, above text, or in the corner of the creative), making it difficult to separate the effect of Logo from other features appearing in the creatives. Therefore Logo did not achieve the expected significance. For Insurance, the result that a Brand Cue was relevant but not a Logo is interesting: either the audience does not associate the brand cue with the brand, or it is seen as a symbol of innovation with more appeal. Text was only significant for Cosmetics, despite it having good impression levels for all the other brands' models. This could be due to the fact that Text was associated with the majority of the impressions (between 62-83%) for the other brands, making it difficult to truly determine the impact of text. It is worth noting that Text in Cosmetics could also be interpreted as a Product, as text would often be on the label (e.g. the scent description).

Person results (Person, Person & Face, Person & Face & Smile, Person & Product, Person & Face & Product) must be read with caution as they are not mutually exclusive. A person smiling was only relevant for the Insurance brand but not in the other brands. Other brands are more Product focused, with more creative impressions featuring products (against Insurance, with only 6% of creative impressions featuring products). It would appear, pending further validation, that when a brand sells 'intangible products' (e.g. insurance products) then smiles and people become more relevant.

For Automotive, Person & Face and Person & Product were the only relevant features. For this brand, 83% of the creative impressions featured a Product and only 56% a Person. Creatives labelled as Product in this brand showed either a whole car, or very close-up images/videos of the inside of cars (with some displaying tech screens). These results would indicate (pending further validation) that adding a person to your product creative generates an uplift in ROI. This provides support for the research discussed in the 'Object Selection' section.

The most efficient features are those which are relevant and have the highest ROI uplifts.

Group	Model	Feature	Rank (1=top): Cosmetic	Rank (1=top): Hospitality	Rank (1=top): Insurance	Rank (1=top): Automotive
Logo, Brand Code & Text	Brand Cue	Brand Cue		7	6	
Logo, Brand Code & Text	Logo	Logo	6	6		
Logo, Brand Code & Text	Text	Text	1			
Logo, Brand Code & Text	Logo Size	Small Logo (0-16%	5	4		
Logo, Brand Code & Text	Logo Size	Medium Logo (16-42%)	4	2		
Logo, Brand Code & Text	Logo Size	Large Logo (42-100%)				
Logo, Brand Code & Text	Logo & Brand Cue	Both		5	2	
Product	Product	Product		3	1	
Product	Product & Logo	Both				
Person	Person	Person	3			
Person	Person & Face	Person w/ Face		1		1
Person	Person & Face & Smile	All			4	
Person	Person & Product	Both			3	2
Person	Person & Face & Product	All	2		5	

Cosmetic · Hospitality · Insurance · Automotive

Table 6: Top performing features by brand

That is, the creatives that display a feature are much more efficient at boosting the ROI than all the other features or none of the features appearing. Table 6 outlines the top-ranking features in terms of ROI uplift for each brand.

The features with the largest uplift across brands (ranked 1) were **Product, Text and Person with Face**; Text was the most efficient in Cosmetics, Person with Face in Hospitality and Automotive and Product in Insurance.

Due to the nature of Text in Cosmetics, this specific feature should be read with caution, however. When looking at the pattern of text along with the other features in Cosmetics, it is found that Text was always accompanied by a Product. For that reason, we cannot say whether Text in isolation drives higher ROIs, but Text with Product was seen to have a positive uplift. A sub-model containing just Product & Text features could be run in order to answer this question.

Excluding Text, the most efficient features were thus **Product and Person with Face. These two features were also shown to be important in combination;** the combination of Person & Product ranked second in both Cosmetics (with Face) and Automotive (without Face).

Brand Cue was also a significant feature in the analysis, ranking sixth and seventh for Insurance and Hospitality, respectively. Similarly, Logo ranked sixth for Hospitality and Cosmetics. When combined, these two features were found to be even more efficient for Insurance and Hospitality, as Logo & Brand Cue was ranked second for Insurance and fifth for Hospitality.

Logo Size was tested in a similar sub-model to just the Logo feature, but the Logo impressions were split out according to size: Small, Medium or Large (determined by the % of the creative which detected the logo occupied). The partner variable was the same as in the Logo sub-model: No Logo. No significant results were found for Large Logo due to the low volume of impressions associated with this feature (<3% for all brands).

For Hospitality, Medium and Small Logo were ranked second and fourth, respectively, while for Cosmetics, Medium and Small Logo were ranked fourth and fifth respectively. Both results indicate that having a bigger Logo is more efficient than having a Small Logo, and in general, having a Logo is better than having no Logo at all.

Many of the results could not be used because they did not meet the minimum impressions threshold (5%), or the uplift was <1 (meaning that a feature appearing was less efficient than not appearing).

This was especially relevant for the Logo, Brand Cue & Text category.

An interpretation of a <1 uplift is that splitting the creative impressions by having a Logo (Yes/No) did improve model performance and achieved significant coefficients to its splits, but the Logo split had a lower ROI than the not having that split. This does not mean that having a Logo in your creatives on Meta will decrease your ROI if you are an insurer, but rather that other features are more efficient than a Logo.

Testing the appearance of a feature vs its partner (e.g. Logo appearing vs not appearing) means that the partnering feature contains all other features. In the Insurance example, in the Logo sub-model, we are testing the appearance of a Logo vs 'everything else' (could be a Product, a Landscape, a Brand Cue, etc.).

Does this mean that all other features are more efficient at driving sales than a Logo? No, there could be one single feature that is much more relevant than the rest and making the Logo partnering split more efficient.

In general, these results should be interpreted with caution and should always be considered in context of the brand and the creatives included in the study.

For example, knowing that the creatives in the Cosmetics brand often included products with labels on them helps in understanding the results for this brand.

Whilst most sub-models showed a positive uplift of the feature appearing, for Text, most sub-models results indicate that adding text to a creative would lead to a lower ROI.

Failing to Define Labels at the Start

Define the set of labels that are going to be studied at the very beginning. Adding more labels in the middle of the study would involve having to go back to time consuming tasks, such as manual labelling and repeated extraction and processing of labels.

Ambiguous Object Definitions

Describe clearly from the start what each label is (e.g. "Person" can be any body part, not just a whole body with a face). This is helpful when the manual labelling is being done as a team, rather than by one person. Furthermore, if using a pre-trained model, ensure that your definition of the object aligns with what is detected by the model. For example, you may define "Car" as just the outside of the car while the OD model is trained to detect both the interior and exterior of cars.

Non-Generalizable Labels

If you are studying two separate sub-brands within one brand, it is advisable to have two separate studies, rather than one. That is, instead of defining objects "Brand A Logo" and "Brand B Logo", it may be better to separate the brands into different streams and have the same object labels for both (e.g. Logo, Brand Cue, Product, and Person). This will ensure that your code is reusable for studies of brands that have different numbers of sub-brands.

Lazy Manual Labelling

Make sure to manually label all objects in a creative. For example, if there are three cars, label all of them, not just one. The manually labelled validation set is the ground truth against which a model's performance is compared. If some objects are missed, you may have strange performance results that indicate that the model may be over-detecting objects.

Trying to be Exhaustive

Avoid testing all open-source resources available, as this can be very time-consuming and not very fruitful. Choose two or three to test and instead spend more time on what you can do to improve their performance on your dataset. This could, for example, be done through hyperparameter tuning (e.g. testing different learning rates, batch sizes, confidence thresholds, etc.) or in the processing of the results (e.g. correcting any text labels inside logos or products).

Lack of Automated Data Errors

Due to the many different data manipulation steps in this project, there are a lot of potential sources for errors. The key is to set up automatic checks at each stage to avoid a trickle-down effect of avoidable errors. For example, removing false positives in face detection, by only 'accepting' a detected face if a person was also detected in the creative, will ensure that the feature time series that is used for MMM does not suddenly have more impressions for faces than for people. Similarly, when doing the feature engineering, employing a simple method of checking that there are no negative values, no missing data, and that the impressions and spend data is consistent across each sub-model will ensure that time is not wasted in the MMM stage from modelling with incorrect data.

Data

Complete dataset: Ensuring you have a complete data set at the start would ensure that you do not need to add ad hoc layers of consolidation. This is easily done by, for example, checking that all the creative ids in the impressions are accounted for in the actual creative files, and vice versa.

Storage

Cloud storage: Due to the volume of creatives included in this study, cloud storage was crucial. Particularly if extracting frames from videos, it is important to account for the significant volume of additional images that need to be stored. This project used Azure Storage.

Object Detection

Labelling Software: For this project, the Azure Machine Learning Studio Data Labelling functionality was used. Having an intuitive software, that can import creatives directly from the cloud storage, and is able to export labels in a familiar format (such as a JSON file) is useful.

External Computation Resources: As the training, validation and final labelling of images are all computationally expensive processes, the use of external computation resources (clusters) is recommended. For pre-processing and feature engineering, individual CPU-enabled single-node clusters are sufficient. For the training, validation, and labelling processes, it is recommended to use GPU-enabled clusters. For this project, Databricks was used as it can connect to Azure storage, facilitates the use of clusters, supports various programming languages, and allows for collaboration on Notebooks.

MMM

Contextual Knowledge: Having a strong understanding of the brand acts as a crucial foundational for every step of this project. It is not only required for making informed decisions regarding which objects should be detected but also vital for defining the features to be measured in the MMM. For example, knowing that products often appear in creatives alongside just a hand raises the question of whether this is the most effective use of a person, or if including the person's face would be more effective; this in turn leads to the creation of features testing products alongside 'face-less' people vs. people with faces. Contextual knowledge can also be gained throughout the project by stopping to analyse the data. For example, checking the distributions of manually labelled objects can give an early indication of performance of custom-trained models (feasibility for successful training and detection) as well as the expected impact for regression models.

Base model: Having strong base models is also key for success in this project. Since the target variable of the sub-models is determined by the contribution of the Meta variables in the base model, a poor base model will directly impact the performance of the sub-model. The quality of the base model will largely depend on the dataset used, so ensuring that sufficient, relevant, and good quality data relating to the baseline, market variations, and marketing activity is crucial.

Programmatic Sub-Modelling: Depending on the number of feature groups, KPIs and sub-brands included in the study, it may be infeasible to run the sub-models on a one-by-one basis. For context, this project had 156 sub-models (13 base models x 12 feature groups). For that reason, it is recommended to create a methodology that allows for the programmatic creation of sub-models.

Data used for this meta-analysis

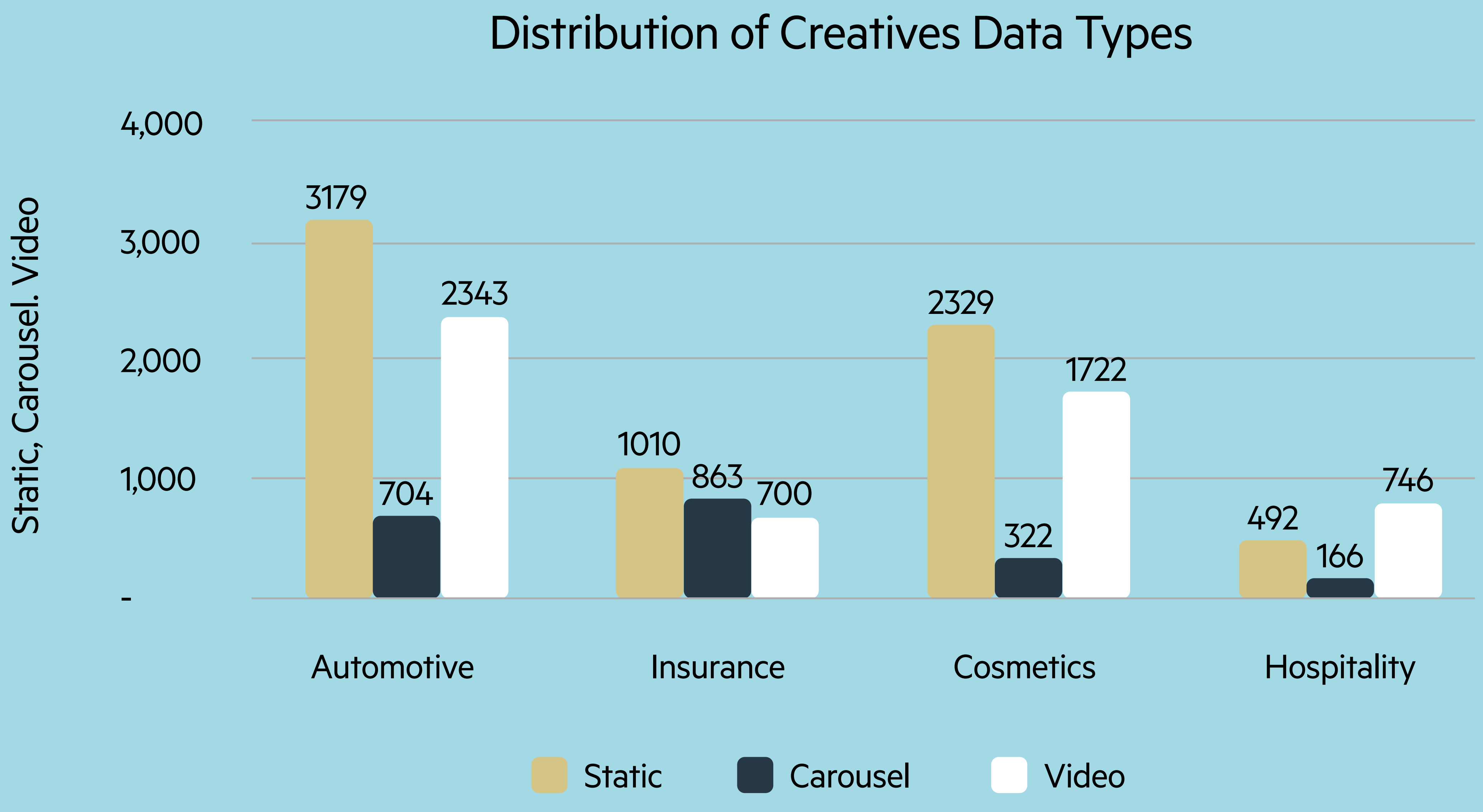


Figure 5. Volume of Creatives by Type and Brand. Total volumes after processing of Videos: Automotive (85,292), Insurance (28,089), Cosmetics (46,656), Hospitality (25,75).

Figure 6: Daily spend and impressions for creatives analysed for the Cosmetics Brand

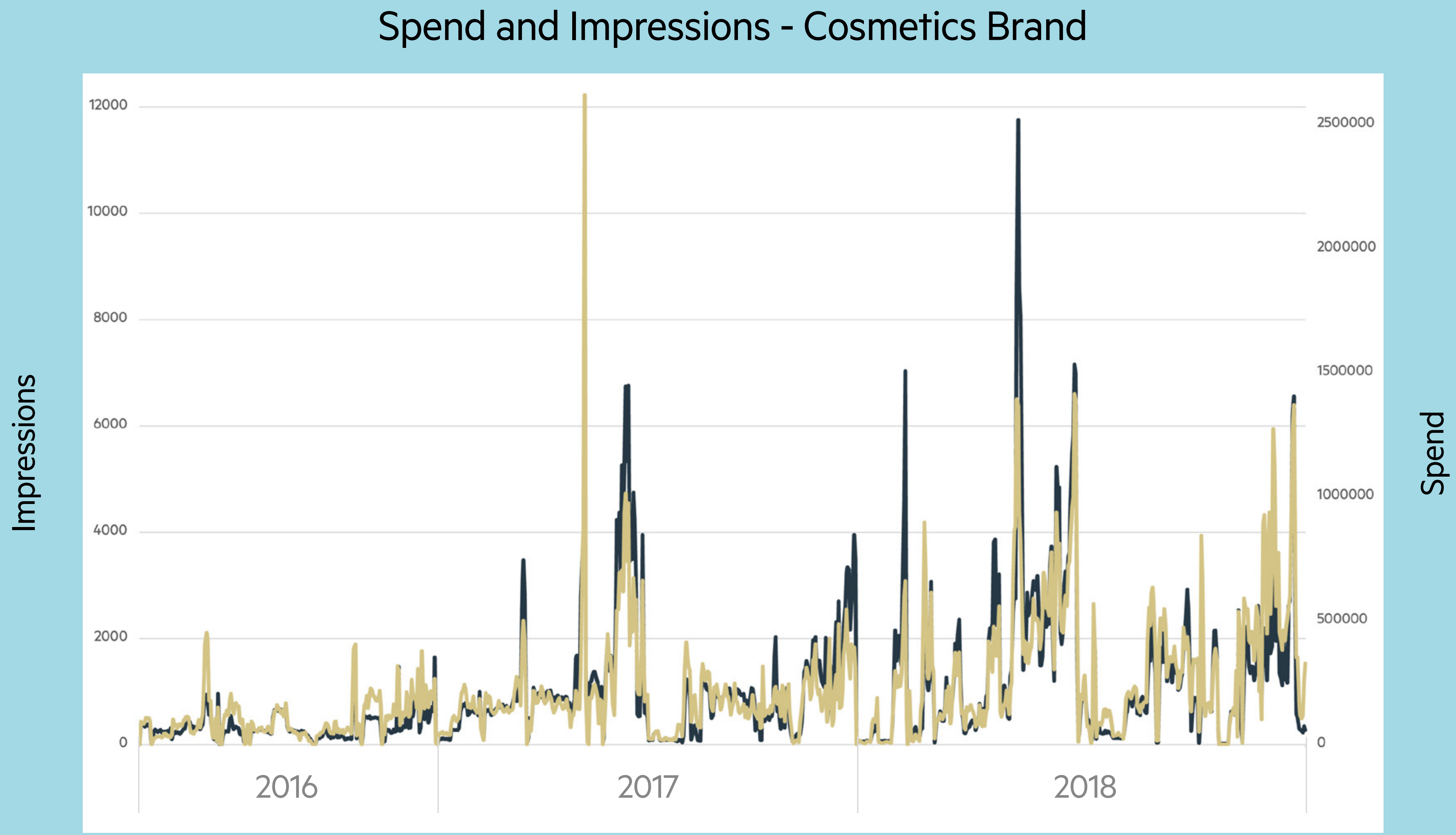


Table 8: MMM Base Models Dataset Details by Brand

Brand	Number of Modelling Datasets	Modelling Period	KPIs
Automotive	4	122-183 weeks	Orders
Insurance	3	45 weeks	Quotes, Web Traffic, Sales
Cosmetics	4	117 weeks	Consideration, Sales
Hospitality	2	74-92 weeks	Sales

Pre-Trained Models

Pre-trained Detectron2 models are trained on the Microsoft COCO (Common Objects in Context) dataset; a large-scale object detection, segmentation, and captioning dataset popularly used for computer vision projects. It contains over 200k labelled images, across 80 object categories (person, car, etc.) and 90 stuff (sky, grass, etc.) categories.

However, for this study, we focused only on the detection of objects relating to the features of interest, detailed in Table 9.

Table 9: COCO Objects Detected in Pre-Trained Models, Present in Our Creatives.

Feature	**COCO Objects Detected**
Hospitality Product (Food & interiors)	Cup, Cake, Bottle, Bowl, Sandwich, Donut, Wine Glass, Orange, Spoon, Chair, Couch, Dining table, Bed Car
Autmotive Product	Car
Insurance Product (Technology)	Laptop, Cell phone, Laptop
Person	Person

Choosing a Detectron2 Pre-Trained Model

Definition Box 1

Accuracy is the percentage of creatives, in the validation dataset, where the type and number of objects detected matches the number of manually objects labelled.

The Confidence Score is the average confidence score for all detected objects in all creatives, in each class.

The performance on Accuracy and Confidence Score (see Definition Box 1) were compared for three models (Figure 8 shows the results):

Model 1: Faster R-CNN R50 FPN 1x

Model 2: Faster R-CNN R101 FPN 3x

Model 3: Faster R-CNN X 101 32x8d FPN 3x

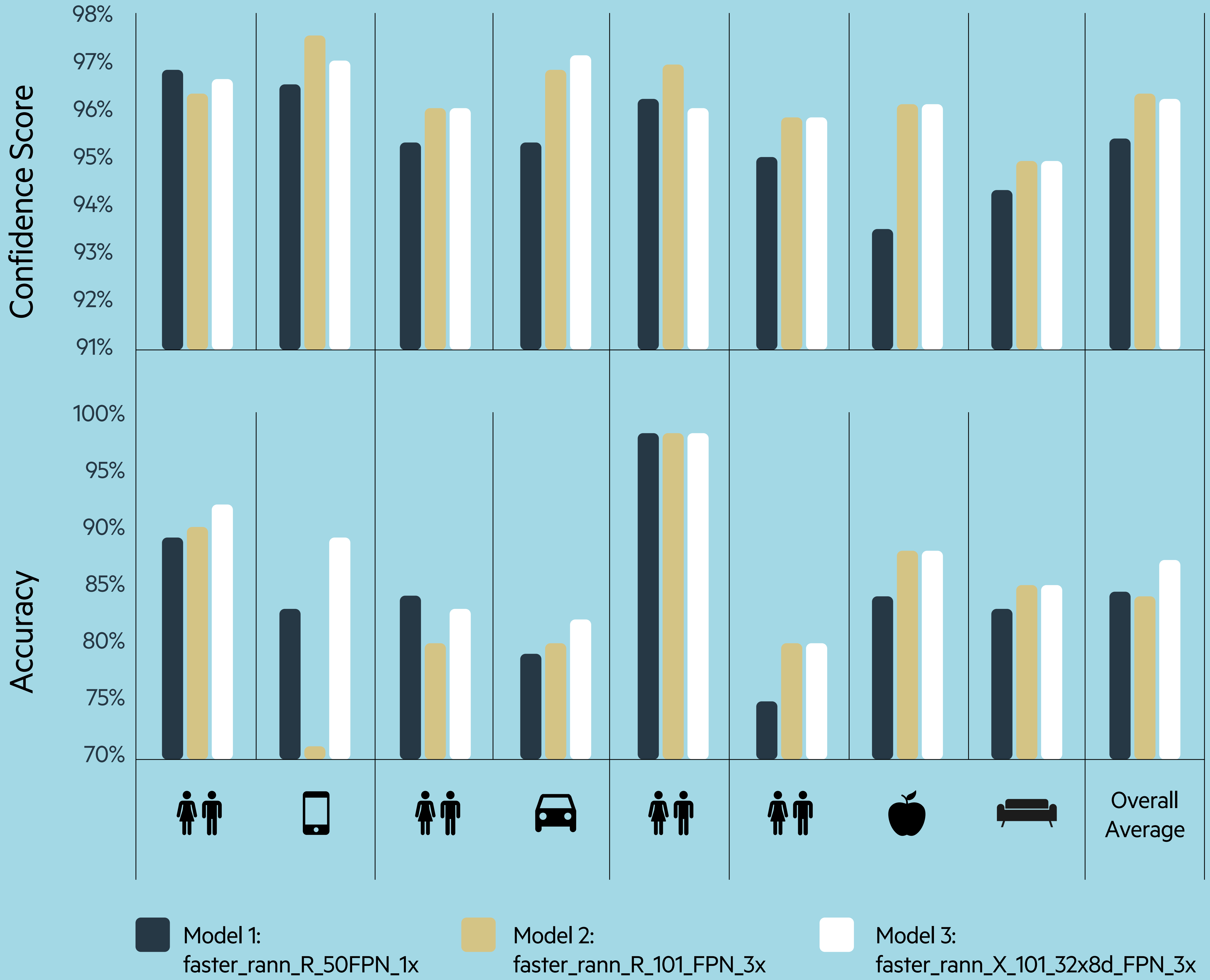
Overall, Model 3 has the highest average Accuracy while Model 2 has the highest average Confidence Score. The number of creatives with correctly identified objects is highest in Model 3.

More in detail, for each object type (figure 8):

- People: All three models have varying performance in detecting people across different brands. Overall, Model 3 has the highest accuracy.
- Technology: Model 2 is most confident in detecting tech, but misses tech in many creatives entirely. Overall, Model 3 has the highest accuracy.
- Cars: Model 3 has the highest average confidence and accuracy in detecting cars.
- Hotel Interiors and Food: Model 2 and Model 3 have a similar performance in detecting food and interior, both higher than Model 1.

Furthermore, Model 3 overcomes challenges faced by the other two models, such as identifying people in close-up shots, identifying cars from interior shots, and identifying technology when in a person’s hand or lying on a table.

Figure 8. Comparison in performance on the validation dataset of three pre-trained Detectron2 models.



Tables 10 and 11 show the performance of Model 3 on the specific Objects of interest. In particular, the Confidence Scores (Table 11) that lead to the high Accuracies in Table 10 range from between 80% to 98%. As we developed our methodology to be applicable regardless of the brand, we needed to choose a Confidence Score Threshold of Acceptance that is high enough to ensure accuracy, but not too high that it would miss objects of interest entirely. Therefore, a threshold of 85% confidence was selected.

Insurance	People	96.6%
	Technology	97.0%
Cosmetic	People	96.0%
Automotive	People	96.0%
	Car	97.1%
Hospitality	People	95.8%
	Food	96.1%
	Interior	94.9%

Table 10. Accuracy of Faster R-CNN X 101 32x8d FPN 3x on Specific Object Types.

Insurance	People	92%
	Technology	89%
Cosmetic	People	98%
Automotive	People	83%
	Car	82%
Hospitality	People	80%
	Food	88%
	Interior	85%

Table 11. Average Confidence of Faster R-CNN X 101 32x8d FPN 3x on Specific Object Types.

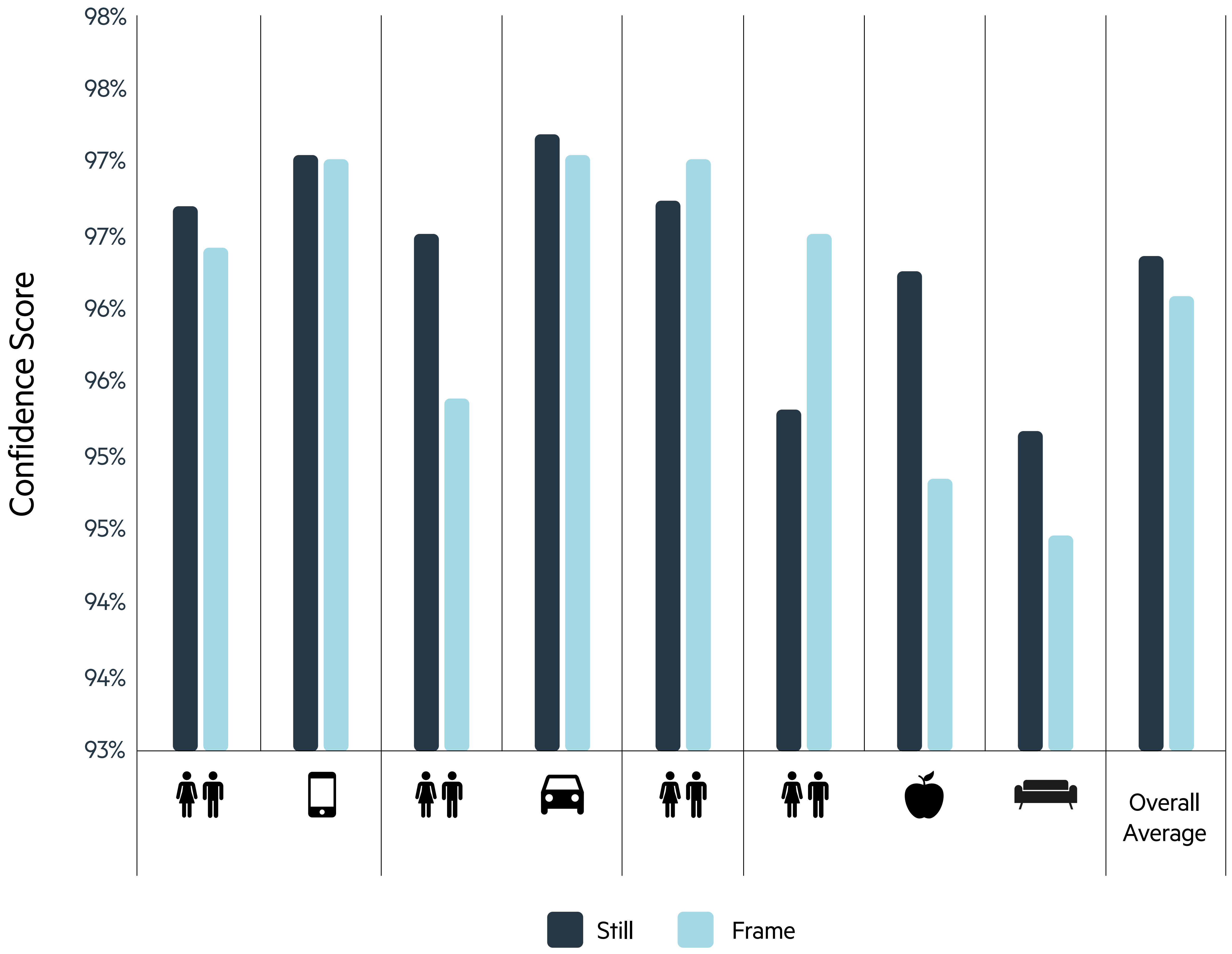
Finally, since our creatives comprise both still images, as well as frames extracted from video files, we wanted to make sure Object Detection was accurate in both types of images.

Figure 9 shows the performance of Faster R-CNN X 101 32x8d FPN 3x (Model 3). The confidence scores do not differ significantly between frames and still images (at most 1.4% points). This means that the model can recognise objects of interest with equal confidence in both types of images.

The differences in performance may be attributed to people in frames more often being close-up; cartoons appear in frames while they are absent in stills; and frames of cars over-index on close-up shots and interior shots.

The total detection time for all brands was 210 hours.

Figure 9. Performance of Faster R-CNN X 101 32x8d FPN 3x on Still vs Frame images.



Detecting Text with Tesseract

Tesseract is an open-source optical character recognition Engine that allows for the recognition of text characters within a digital image. It is an open-source resource, originally developed by Hewlett-Packard, and now managed by Google [3]. This package does not have any parameters to optimise, but, as is exposed here, the basic performance of this resource can be improved with a combination of image processing and detected text correction.

In this work, Tesseract was used to detect Text in all images, in a process outlined in Figure 13. Performance of this detection tool was done using Confusion Matrices, to gain insight not only into the Accuracy, but also other metrics such as the True Positive and True Negative Rates.

To start with, the functionality was used on the original images, to measure a baseline for performance on our sets of images. From this first step, we derived the following learnings:

- For general Text (both non-promotional Text and Promotional Text), the Accuracy is 69%, the True Positive Rate is around 65%, the False Positive Rate is about 35%
- Tesseract does not recognise symbols such as % (which may indicate promotional text) accurately.
- Tesseract does not perform well on images that have a busy background
- Tesseracts does not recognise slanted Text.

Figure 10 Three stage process by which the basic performance (accuracy) of Tesseract on the original images was improved by up to 28% points.

		Insurance	Automotive	Retail	Hospitality	Enhancing Methodology
Step 3	Optimized Accuracy	97%	93%	86%	81%	
		+19% ↑	+16% ↑	+27% ↑	+12% ↑	Restricting text detection to outside of logo or product bounding boxes.
Step 2	Optimized Accuracy	78%	78%	59%	67%	
		+9% ↑				Image processing
Step 1	Proof of Concept	69%				

Building on those learnings, we implemented a pipeline, outlined in Figure 11, that would help us improve the performance from baseline, by up to 28 percentage points on Accuracy (Figure 10):

Figure 11. Pipeline for pre-processing images before Optical Character Recognition Models, and correcting detected text.

Methodology

For each image:

1. Image Pre-processing
2. Detect Text
3. Delete Unwanted Text
    a. If inside bounding box of Logo or Product
    b. If shorter (in number of characters) than a defined threshold (which depends on the brand)
    c. If has key words (e.g. the Brand name)
    d. If it is blank or special characters (ASCII code)
4. Go to the next transformation and repeat the process.

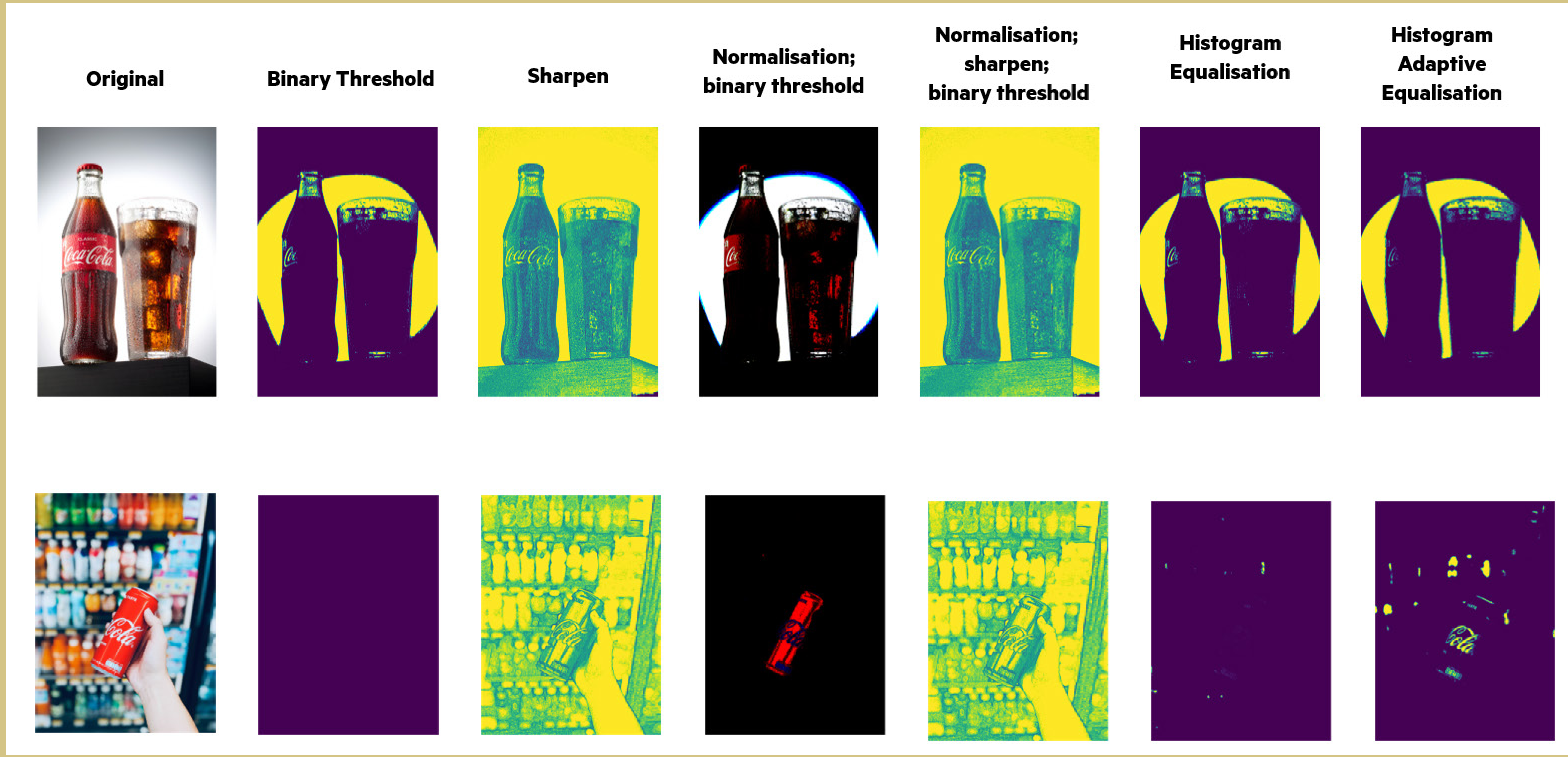
For Step 1, pre-processing methods included the following, as well as combinations of those (illustrated in Figure 12):

Binary Threshold, Sharpen, Normalization, Histogram Equalization, Histogram Adaptive Equalization.

As can be imagined, a processing method that enhances the performance of the detection method for one image, might decrease it for different image. Therefore, to ensure we obtained the highest accuracy for all images, we applied all pre-processing methods to all images, running Tesseract on each modified image, and keeping track of the text detected in each iteration.

Step 2, we use Tesseract to detect text in the pre-processed images.

Figure 12. Effect of various image processing on the original. Images for illustrative purposes.

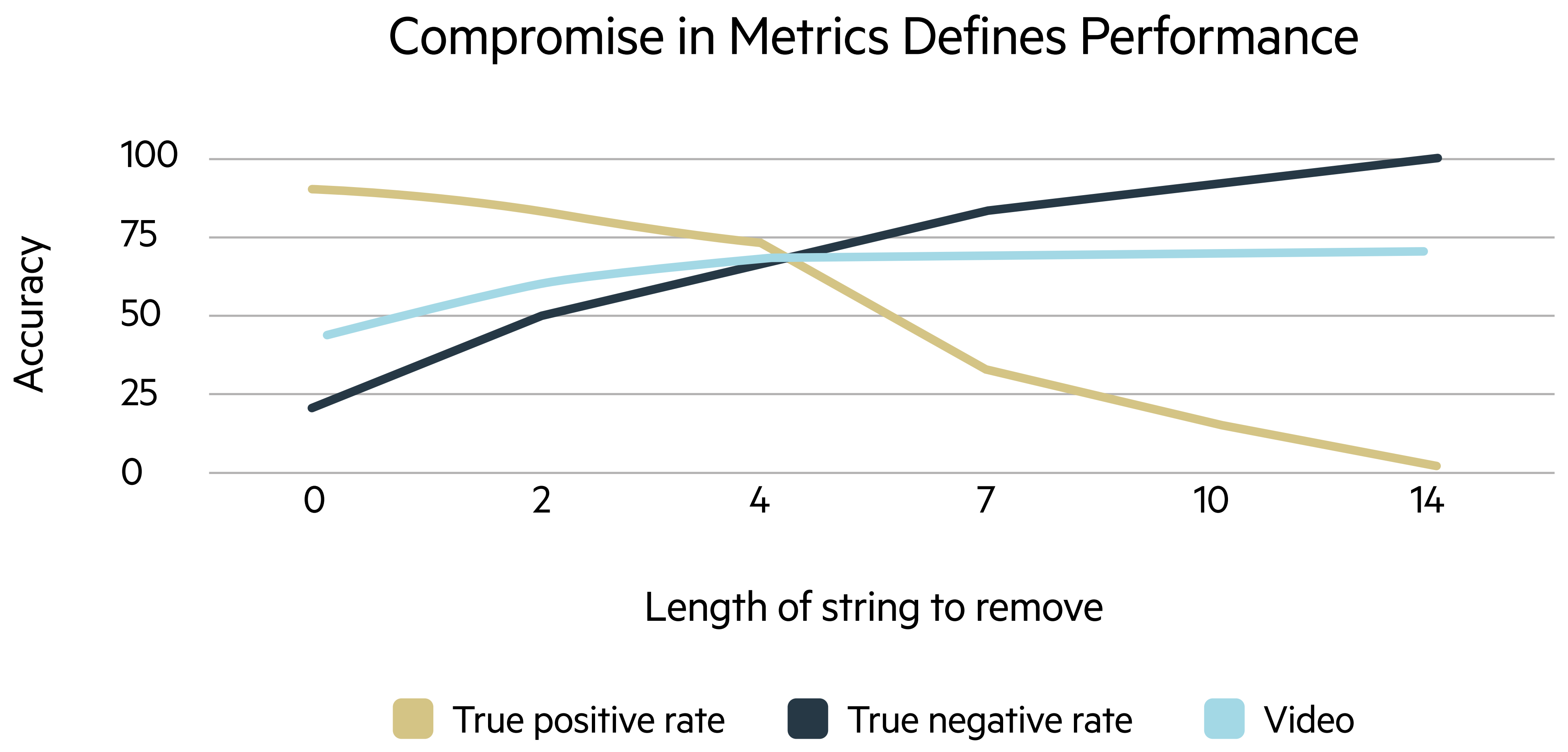


Step 3 consisted of removing any "incorrect" text, such as Text inside the bounding boxes of Logo or Product objects.

If the text was shorter than a threshold length (which depended on the brand), if it consisted of keywords, such as the brand name, or if it had ASCII characters. Figure 13 shows that with the current methodology (Figure 11), the accuracy cannot be improved any further by changing the length of text that is considered to be "true" Text. Furthermore, the image shows the threshold for the length of strings to accept as true Text, is the value that optimizes both the True Positive Rate and the True Negative Rate (i.e. four characters long in that particular case).

The performance of Tesseract may be improved further by adding another step to the pipeline in Figure 11, where the image is rotated several degrees, and after each rotation, Text is detected. Due to time constraint issues, and the fact that we had already achieved a significant improvement above baseline and reached an accuracy of high 90s for some of the brands, we did not implement this step.

Figure 13. Graph showing the trade-off between the True negative Rate and the True Positive Rate helped us choose a threshold for the length of string to accept. Since Accuracy does not improve further after length 4, but this value does optimise the other two metrics, it is the threshold chosen.



Custom-Trained Model

The Faster R-CNN model, Faster R-CNN X 101 32x8d FPN 3x, was also used for detecting brand-specific objects which were not included in the pre-trained library (Logo, Brand Cue and cosmetic Products), as well as Faces and Smiles. Initially, Faces and Smiles were detected by pre-trained algorithms but due to the poor performance, it was decided to use a custom algorithm for these as well.

One model was trained per object per brand using the manually labelled training sets. For detecting Faces and Smiles, the training set consisted of creatives from all four brands, but each model was developed separately for Face and Smile per brand. In total there were, thus, 19 custom models. The validation set was used to tune the hyperparameters of each model, with accuracy as the main metric. The final models were then used to detect objects in the unlabelled images for all four brands.

Features engineered for MMM

Table 13. Features engineered, along with their groupings.

No.	Feature Group	Features Included	Partner Feature
1	Logo	Logo	No Logo
2	Brand Cue	Brand Cue	No Brand Cue
3	Person	Person	No Person
4	Person & Face	Person & Face; Person only	No Person
5	Person, Face & Smile	Person & Face & Smile; Person & Face only; Person only	No Person
6	Product	Product	No Product
7	Text	Text	No Text
8	Logo & Product	Logo & Product; Logo Only; Product Only	No Logo or Product
9	Logo Size	Small Logo; Medium Logo; Large Logo	No Logo
10	Logo & Brand Cue	Logo & Brand Cue; Logo Only; Brand Cue Only	No Logo or Brand Cue
11	Person & Product	Person & Product; Person only; Product Only	No Person or Product
12	Person, Face & Product	Person & Face & Product; Person & Face Only; Person & Product Only	No Person or Product

The Authors

Matt Andrew
UK MD & Partner

Matt likes to be hands-on, steering clients to new ways of looking at and implementing marketing effectiveness, delivering optimal value and highly predictive models from a combination of methodologies.

He began his career with Colgate Palmolive, where as a Senior Brand Manager he was responsible for multi-million-pound budgets, media planning and strategy, product strategy and market research, giving him unique client-side insight. Matt then went on to cut his teeth in analytics at former Dunhumby founders' fast-growing, innovative start-up Starcount, before joining Ekimetrics and building the UK business.

matt.andrew@ekimetrics.com

Marina Bermejo
Manager, Ekimetrics

Marina is an experienced consultant, solving complex problems though analytics across industries, including Media & Entertainment, Insurance, Retail, Hospitality and CPG. She co-leads the UK MMO Solution Team, fuelling innovation globally.

Marina loves to understand the contextual nuance of her clients, find the right analytical approach and translate the outputs into actionable insights that get results.

She started her career as an Analyst in Madrid, before moving to London and Gain Theory, where she was rapidly promoted over the four years she was there. She has a BSc in Economics and a Masters in Econometrics, as well as having completed a course in Business Strategy and Consulting at Imperial College London's Business School.

Monica Brondholt Sorensen
Senior Data Science Consultant, Ekimetrics

Monica is an exceptional academic, with a Dean's List MSc in Business Analytics from Imperial College Business School – one of the most selective master's in the world – and a history in academic research at the University of Toronto. Her first degree was in Economics, Mathematics and Human Geography. She marries this with the practicalities of business applications of analytics to deliver actionable insights that transform disciplines from pricing, new product development and digital marketing to partnerships and employee enrichment.

Monica was promoted within her first nine months at Ekimetrics, having begun her commercial career in business intelligence.

Karin Sasaki
Senior Consultant, Ekitmetrics

Karin brings over five years of experience in modelling and data analysis across a range of industry and academic settings to Ekimetrics. She has worked in fields as diverse as molecular biology systems to operational research.

Following her PhD in Mathematics from Imperial College London, Karin held a number of post doctorate research fellow positions, as well as supporting and teaching other academics, before moving into marketing applications at Kantar, Capita and now Ekimetrics. Her breadth of skills means she's able to bring fresh thinking and an accessible style to her consultancy.

Leveraging data effectively is complex and challenging. We're here to get to the crux of how it can help your business.

If you're in need of a straightforward conversation, contact Matt today.

About Ekimetrics

Global data science leader, Ekimetrics, helps clients optimize their marketing, customer engagement and operational efficiency, as well as advising how to upgrade their products and services.

Founded in 2006, Ekimetrics has a presence in more than 50 countries, generating over EUR 1 billion in profits for its clients since 2006.

ekimetrics.com


WellChild
Giving seriously ill children the best chance to thrive. At home.
Text CHELT 10 to 70085 to donate £10*
I AM JD
SHOP 24/7 ONLINE NOW
ONLY AT JD
SAMSUNG
GAP
Boots
GAP
Ekimetrics. × Meta