KONICA MINOLTA

THE FIRST INTERNATIONAL CONFERENCE ON IOT-AI

DECIPHERING BRAND IDENTITY FROM PACKAGE:
VISUAL FEATURE ANALYSIS THROUGH CONVOLUTIONAL NEURAL NETWORKS

A. SHIMOJO and S. URATANI

Presenter: Asaya SHIMOJO

Konica Minolta, Inc.

asaya.shimojo@konicaminolta.com

June 30 to July 04, 2024

@Porto, Portugal

IARIA

Giving Shape to Ideas

PRESENTER

KONICA MINOLTA

Asaya SHIMOJO

asaya.shimojo@konicaminolta.com

Professional Experience:

- Researcher at Konica Minolta, Inc.
  (Data Science)
- Ph.D. candidate at Nagoya University, Japan
  (Cognitive Science)

INTRODUCTION

Problem

Many cases of product packaging renewal resulting in a loss of that product's identity and a drop in sales

→ Because design elements that contribute to the identity of the product brand have been changed and the identification rate has declined

INTRODUCTION

KONICA MINOLTA

[Previous study]

- Recent advancements in AI technology have enabled the creation of CNN models that can learn and classify package images of specific brands.
- Notably, brand identification using CNN has been validated in studies involving brand logos and fashion show runway photos [1][2]

- Grad-CAM (Gradient-weighted Class Activation Mapping) is a technique that visualizes as a heat map the image regions that the model focuses on during classification. This facilitates model interpretation [3][4]
- However, it does not directly indicate the areas of attention that a human consumer pays attention to when identifying a package; comparisons of AI models' regions of attention with human visual attention have not been well validated [5]

INTRODUCTION

KONICA MINOLTA

Experiment 1

Multi-class Classification Model based on VGG16

1. Data Collection and Preprocessing
2. Model Training and Validation
3. Performance Evaluation

Model Construction

Experiment 2

The Alignment between Human and the Model Attention Regions

1. Psychological Experiment for Collecting Eye-tracking Data
2. Heatmap Generations
3. Comparison between Human and AI Visual Attention

Model = Human ?

Experiment 3

Filter Analysis and Ablation Study

1. Extracting and Analyzing Filters
2. Ablation Study
3. Analyzing the degree of Filter's Contribution on Model Decisions

Contribution areas

EXPERIMENT 1:
What does the model look for in a package to make a decision?

KONICA MINOLTA

[Model construction]

Build a model that, given an input package image, classifies with high accuracy which of the brands A~E it corresponds to.

Data Collection and Preprocessing

Collect 6,000 package images from Google for 5 stationery brands.

※ Only those works for which the copyright has expired.
※ The brand name were masked.

e.g., Brand A's Packages

120 packages

× 5 brands = 6000 packages

- Brand A: Luxury brand
- Brand B: Luxury brand
- Brand C: Mass-market brand
- Brand D: Mass-market brand
- Brand E: lesser-known brand

Model Training and Validation

- **Model**:
  - VGG16 pre-trained on ImageNet
  - Output size of the final layer adjusted to 5 to classify brands A through E
- **Dataset**:
  - 70% of the dataset for training
  - 30% of the dataset for validation
- **Training and Validation**:
  - Conducted 2,000 times
  - The data was shuffled each time
  - With a batch size of 32 and 10 epochs

→ a validation accuracy of 91%

EXPERIMENT 1

KONICA MINOLTA

[Grad-CAM]

To identify CNN visual attention when recognizing the brand of a package

The model used for classification is finally converted into 4,096 features by convolving in 5 layers, so it is difficult for the human eye to know what is contributing to the classification result, The **Grad-CAM** method makes it somewhat easier for humans to recognize.

Original Image

Grad-CAM

EXPERIMENT 2

KONICA MINOLTA

Experiment 1

Multi-class Classification Model based on VGG16

1. Data Collection and Preprocessing
2. Model Training and Validation
3. Performance Evaluation

Model Construction

Experiment 2

The Alignment between Human and the Model Attention Regions

1. Psychological Experiment for Collecting Eye-tracking Data
2. Heatmap Generations
3. Comparison between Human and AI Visual Attention

Model = Human ?

Experiment 3

Filter Analysis and Ablation Study

1. Extracting and Analyzing Filters
2. Ablation Study
3. Analyzing the degree of Filter's Contribution on Model Decisions

Contribution areas

EXPERIMENT 2

KONICA MINOLTA

[Human eye-tracking]

To identify human visual attention when recognizing the brand of a package

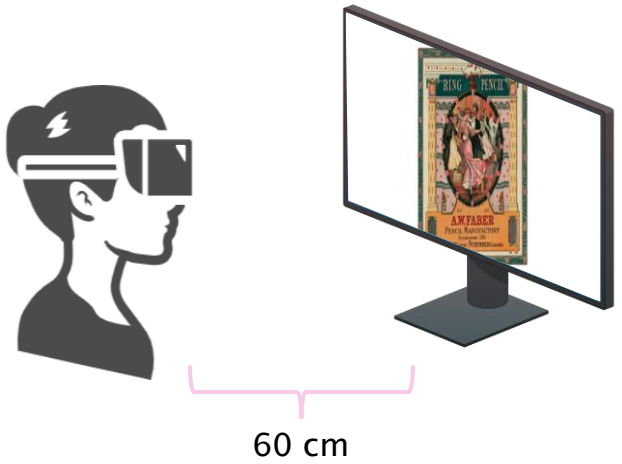
Participants

6 Adults 2 female and 4 males (39.8 ± 7.9 years) having normal visual acuity.

Procedure

Looking at a package, participants:

1. Identified which brand from A to E the package belonged to
2. Rated their confidence level in that classification (0–100%)



× 10 times
for learning each brand's VIs

× 20 times
for classifying packages

Eye movements were recorded during only the classification process.

EXPERIMENT 2

KONICA MINOLTA

Eye-tracking

Grad-CAM

Jaccard Index	Interpretation
0.00	no overlap
0.01 – 0.25	Slight similarity
0.26 – 0.50	Moderate similarity
0.51 – 0.75	High similarity
0.76 – 1.00	almost identical

Average heatmap agreement between the two exceeds 0.32.

→ Attention areas of humans and models were found to match to some extent

EXPERIMENT 3

KONICA MINOLTA

Experiment 1

Multi-class Classification Model based on VGG16

1. Data Collection and Preprocessing
2. Model Training and Validation
3. Performance Evaluation

Model Construction

Experiment 2

The Alignment between Human and the Model Attention Regions

1. Psychological Experiment for Collecting Eye-tracking Data
2. Heatmap Generations
3. Comparison between Human and AI Visual Attention

Model = Human ?

Experiment 3

Filter Analysis and Ablation Study

1. Extracting and Analyzing Filters
2. Ablation Study
3. Analyzing the degree of Filter's Contribution on Model Decisions

Contribution areas

EXPERIMENT 3

KONICA MINOLTA

[The aim of Experiment 3]

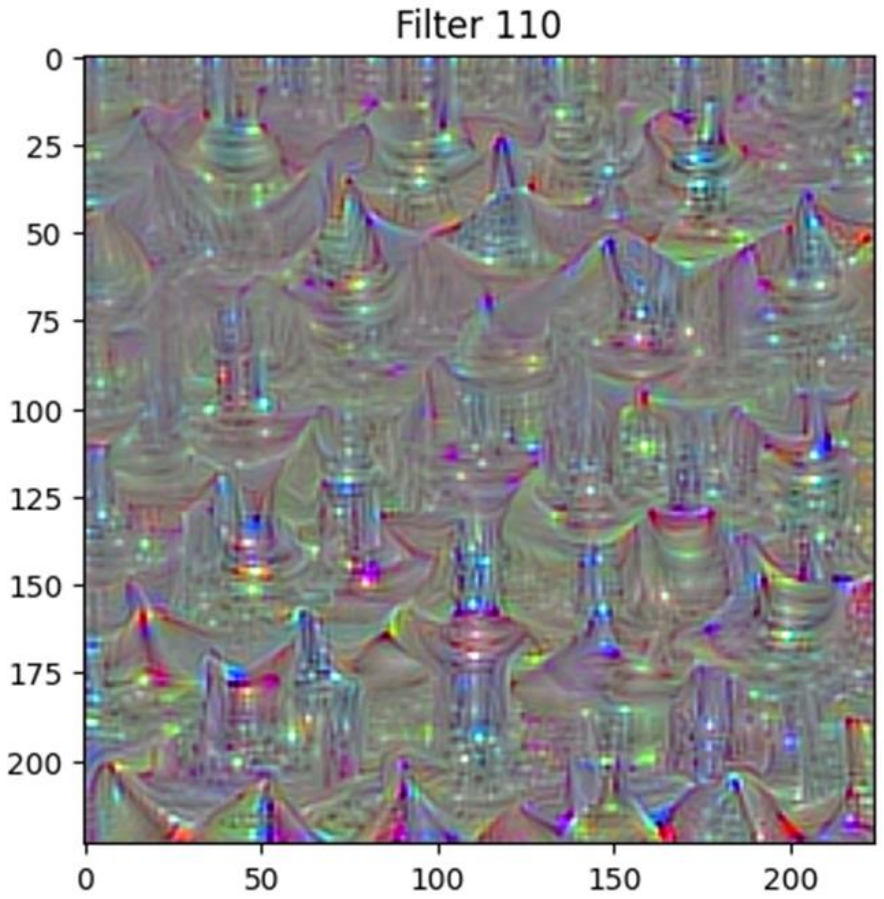
To analyze the visual information processing of the CNN model and identify design elements that contribute to Brand A's brand identity

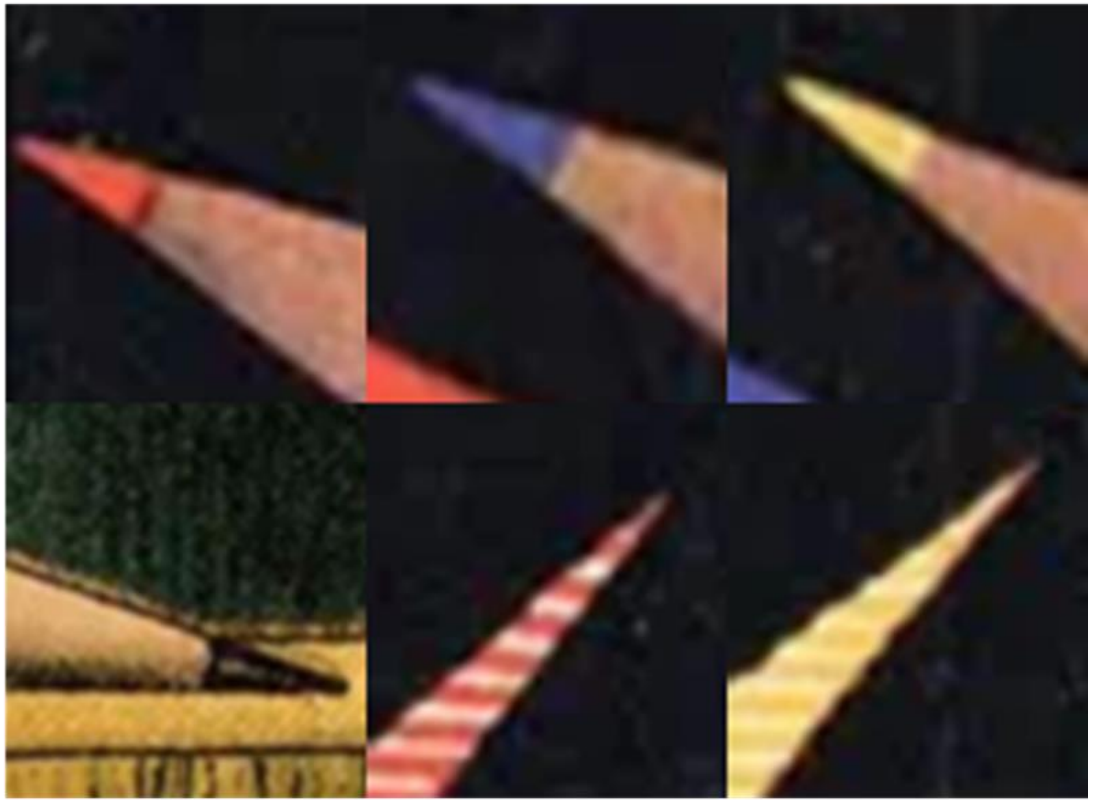
Procedure

1. **Filter Visualization**:
   The filters in the middle layers of the CNN model were visualized to identify the visual elements the model focuses on when recognizing Brand A.

2. **Ablation Study**:
   An ablation study was conducted to assess the contribution of each filter to the model's performance. This involved systematically disabling individual filters and observing the impact on the model's classification accuracy.

EXPERIMENT 3





GENERAL DISCUSSION

KONICA MINOLTA

[Findings]

- We confirm that the machine learning model can identify important design elements in brand recognition.
- In addition, visualization using Grad-CAM showed partial agreement between the visual recognition of humans and machine learning models.
- This will contribute to the quantitative management of brand identity and the development of effective packaging design strategies.

[Future works]

- However, while Grad-CAM is an effective method of highlighting important areas, it tends to be biased toward visually prominent areas.
- Further experimentation and expansion of the data set are needed to generalize the results.


KONICA MINOLTA