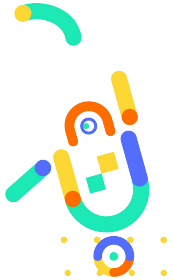
GDG Hanoi

Google I/O 18 Extended Hanoi

Mobile and Server Architecture for Fashion Search System

Mr.. Pham Van Toan

AI Leader Framgia Inc



Who is this guy?

- **Fullname**: Pham Van Toan
- **Job**:
  - **Artificial Intelligence Leader** in Framgia
  - **Technical Advisor** in IK-Home JSC
- **Experiment:**
  - Server architecture for IoT, AI application
  - Computer Vision
  - Natural Language Processing
  - Internet of Things
  - Speech Analysis

Agenda

Deep learning techniques in **Mobile & Server Architecture** for street-to-shop fashion search

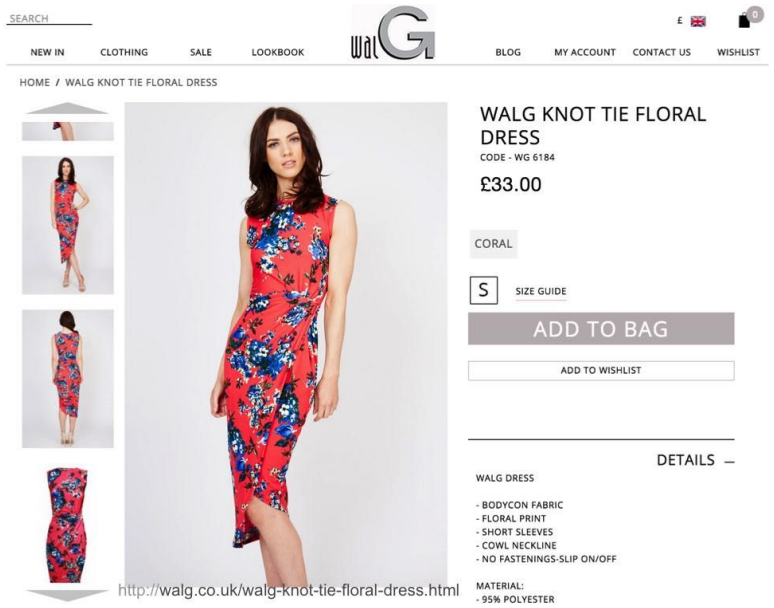
1. **Fashion Search - new trend in online shopping**
2. **Deep learning techniques in Fashion Search**
3. **The power of mobile in deploying deep learning system**
4. **Server architecture in deployment and demo.**

Information Overload

Fashion search in real world

SEE IT

SNAP IT



BUY IT!

Objects detection

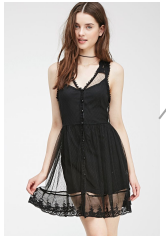
SSD Mobilenet model

Deep ranking model

Query image

Positive image

Negative image

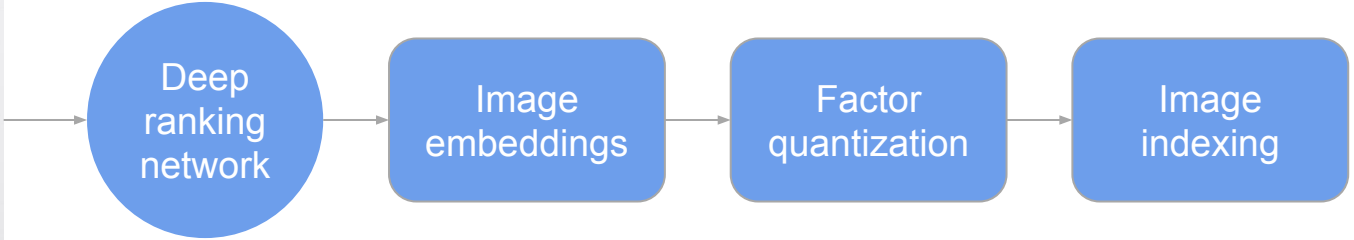


Deep ranking network

Image embeddings

Triplet loss

Image indexing



AI/ML on Mobile advantages

Offline

Running on mobile makes it possible to deliver very interactive applications

Interactive hardware

Solves the problem of cloud services

Save Cloud

Avoid wasting bandwidth, CPU, RAM ...

Smart Apps

Upgrade yourself

#Speech Recognition, #Image Recognition, #Gesture Recognition, #Text Recognition, #Object Detection, #Character Recognition, #Translation ....

Developer Tools

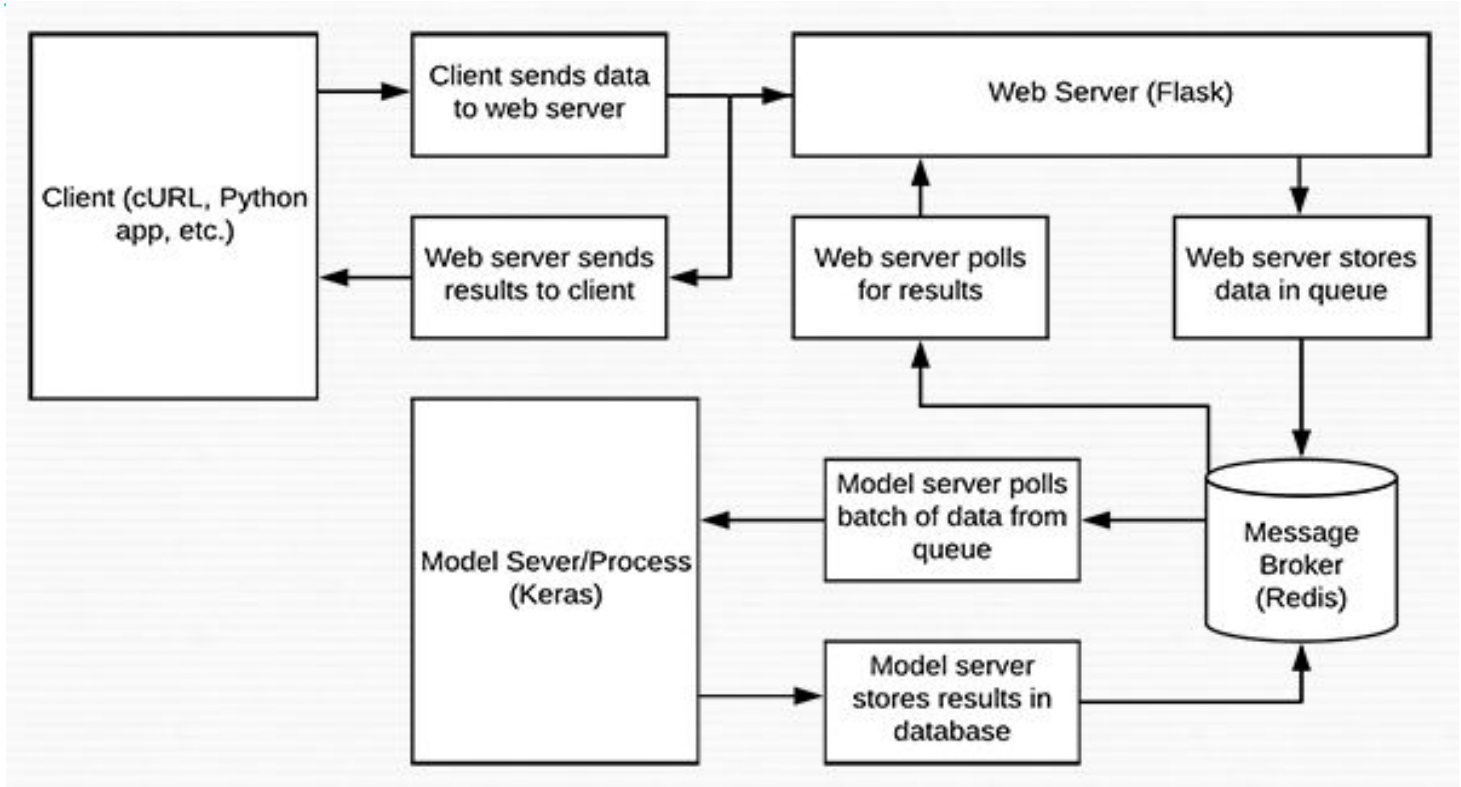
Framework

- Native Framework

  Android: Neural Networks API

  iOS: Core ML
- Firebase ML Kit (TensorFlow Lite)
- TensorFlow Mobile
- Caffe2 (PyTorch)

Deep Learning Web Server



Our Server Architecture

Deep learning framework
Tensorflow, Keras, Scikit-learn, and Numpy

Caching
Redis for publish / subscribe method and in-memory data structure store

Full text search
Elasticsearch for indexing the decoded embedding

Restful API
Django, Node JS for create the RESTful API endpoint

Web Server
Nginx for HTTP server and reverse proxy

Web Client
HTML5, CSS3, JQuery

Deployment
Gunicorn, Supervisor, Rockettee

Server hardware
RAM: 8GB
CPU: Intel® Core™ i5-7500 CPU @ 3.40GHz
GPU: None

Thank you!