



# Visual Search and Recognition at ViSenze

Guangda LI  
CTO and Co-founder of ViSenze

NVIDIA AI CONFERENCE

# Agenda

- About ViSenze
- Deep learning behind ViSenze's Solution
- ViSenze's practices to accelerate the R&D efficiency using deep learning

NVIDIA AI CONFERENCE

A hand holding a smartphone in the foreground, displaying a street scene on its screen. The background is a blurred city street at night with bokeh lights. The text 'About ViSenze' is overlaid in white on a dark horizontal band.

# About ViSenze

NVIDIA AI CONFERENCE

ViSENZE™

## About Us

ViSenze is a global visual search provider, catering to the \$1.67 trillion e-commerce and retail markets. Built for scale and performance, ViSenze processes over 100 million images every day and is used by some of the largest Internet companies like Rakuten, ASOS and UNIQLO.

ViSenze has been recognized amongst:

- Best AI Product in Retail by CognitionX London
- Top 5 deep learning companies by VentureBeat
- Top 20 artificial intelligence companies by DataMation
- Top 40 global Breakthrough Brands by InterBrand



[link](#)

FORTUNE

Forbes

Bloomberg

TNW  
THE NEXT WEB

FORRESTER

recode

FAST COMPANY

VentureBeat

THE BUSINESS TIMES

THE STRAITS TIMES

Interbrand

TC TechCrunch

BUSINESS  
INSIDER

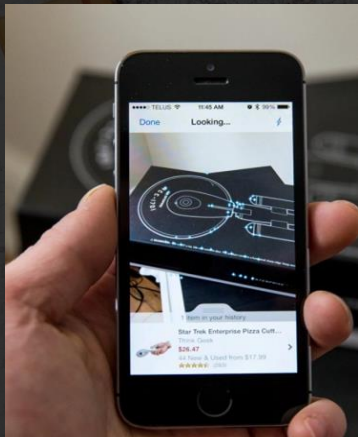
DNA  
DIGITAL NEWS ASIA  
Your Eye on the Tech Ecosystem

TECHINASIA

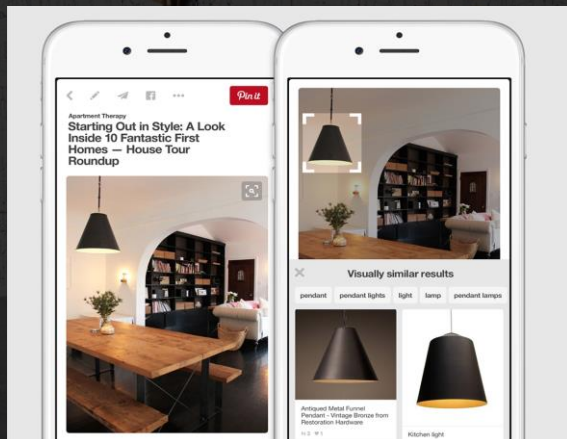
VISENZE™

# Growth of visual content and how major players are driving this shift

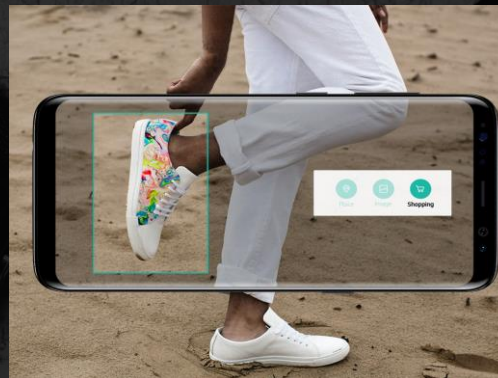
## Amazon Visual Search



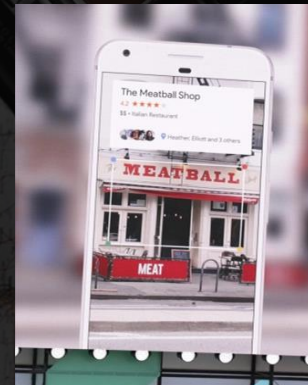
## Pinterest Lens



## Samsung Bixby Vision



## Google Lens



3,000,000,000 *images uploaded online per day, 2015*

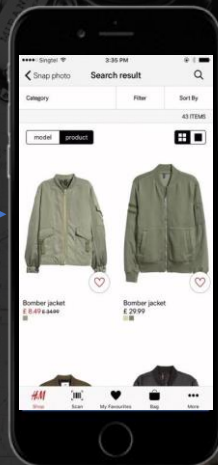
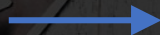
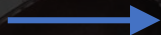
250,000,000 *visual Searches per month, on Pinterest – May 2017*

360,000,000 *Visual Searches per month, on Taobao - Aug 2017*

# Search Solutions and Use Cases

## Visual Search

Enhance search and discovery



## Search by Image

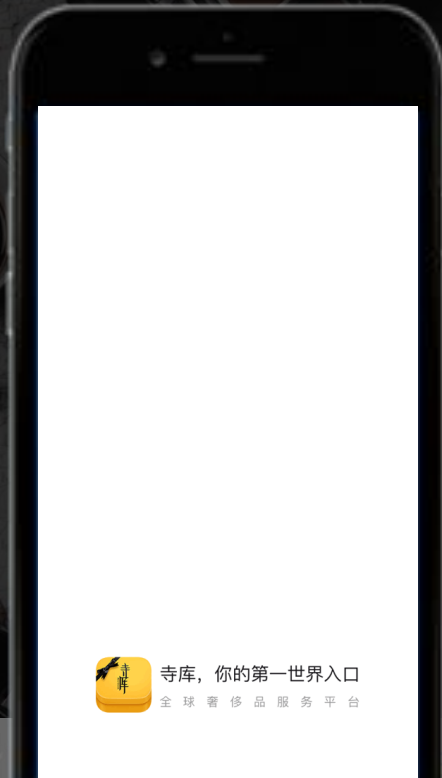
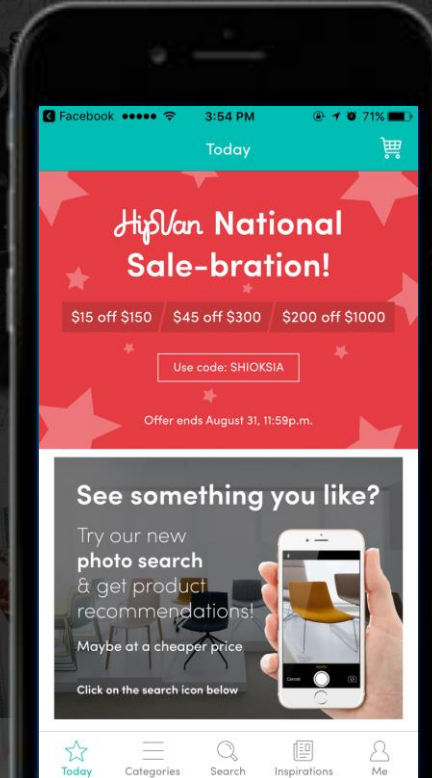
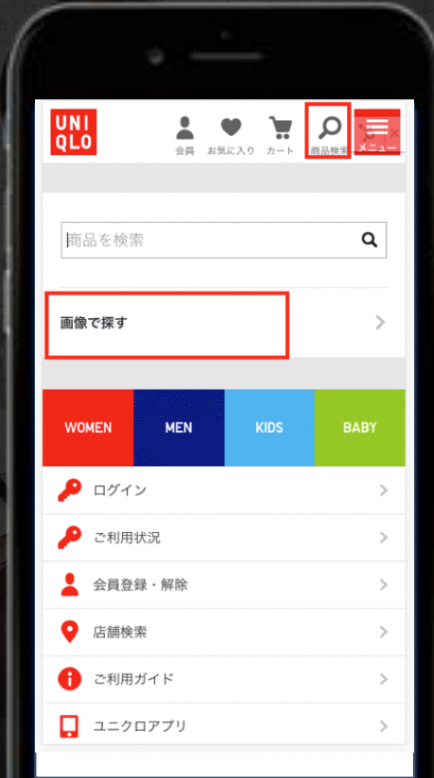
Enable shoppers to search using an image or screenshot

## Search by Color

Color palettes allow users to broaden their experience to new product categories

Image Search is now a common feature on leading retailer apps

# H&M, ASOS and UNIQLO Visual Search powered by ViSenze



# Recommendation Solutions and Use Cases

## Visual Recommendation and Discovery

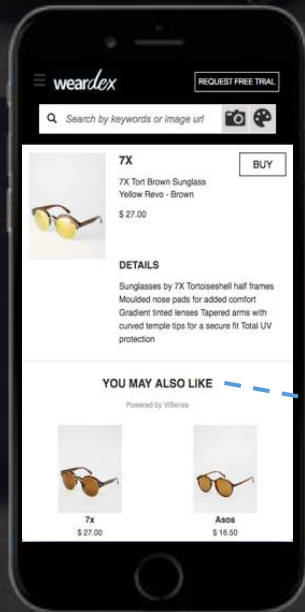
Improve discoverability and conversions

## You May Also Like

Recommendations based on what shoppers are viewing

## Out of Stock Alternatives

Suggest visually similar products when shopper is viewing an out of stock item



**YOU MAY ALSO LIKE**

VISUAL RECOGNITION POWERED RECOMMENDATIONS



# How It Works : Product recommendation

The screenshot shows the Interpark website homepage. At the top, there's a navigation bar with categories like '쇼핑' (Shopping), '도서' (Books), '티켓' (Tickets), and '투어' (Tours). Below this, there's a search bar and a main content area with several product recommendations. On the left, there's a 'new' badge for '우수회원 혜택안내' (Excellent Member Benefit Guide). The main content area features a grid of products including 'World History Effort 1', 'I told you to wash rice an...', and 'eBook! Why? 시리즈'. At the bottom, there's a section for '이벤트로 참여하고 포인트로 한 번 더 할인!' (Participate in events and get an extra discount with points!).

The screenshot shows the ASOS website homepage. At the top, there's a navigation bar with 'WOMEN', 'MEN', and 'OUTLET' categories. Below this, there's a search bar and a main content area with a large fashion advertisement. The advertisement features a man and a woman in casual clothing, with the text 'This is ASOS YOUR ONE-STOP FASHION DESTINATION'. Below the advertisement, there's a 'FREE DELIVERY\* WORLDWIDE' banner. At the bottom, there are two 'SHOP' buttons: 'ASOS OWN LABEL' and 'ASOS BRANDS'.

# Tagging Solutions and Use Cases

## Automated Tagging

Improve discoverability and conversions

Color

Pattern

Fashion Attributes\*

Custom Attributes\*\*

## Catalogue Management

Tag entire image libraries based on visual attributes for better search results

## Image Filtering and Moderation

Tag entire image libraries based on visual attributes

\*Q3 2017 Availability  
\*\* Subject to requirements analysis

# How It Works: Tagging (Demo Video)

**weardex** | PRODUCT TAGGING ▾

REQUEST FREE TRIAL Go to Admin

## Product Tagging Playground

Upload image or try out the demo images below.

Tag by image URL or image upload

**Tagging Result**

**Attributes** Powered by ViSenze

**FASHION**

1

category:	[jumpsuit]	0.658807
neckline:	[o_neck]	0.372548
product_color:	[green]	0.322762
product_pattern:	[other]	0.411989
sleeve_length:	[sleeveless]	0.647277
sleeve_style:	[capped]	0.353547

**OVERALL IMAGE QUALITY**

[no text]	0.9653041
[model]	0.9999987
[no collage]	0.9504868
[no mosaic]	0.9999988
[no detail]	1.0000000

## Improving User Engagement Through Artificial Intelligence

**30%**

*higher conversions  
on image search  
over text based  
search*

**50%**

*higher CTR of  
shoppers who click  
on visually similar  
products*

**5x**

*higher conversion  
rates for shoppers  
clicking on visually  
similar products*

**160%**

*increase in  
engagement for  
shoppers who used  
find similar*

**A.I. powered visual search and recognition  
solutions improve engagement and  
conversions**

# ViSenze Enterprise Ready Solutions



Scalable

Scales during traffic spikes without compromising user experience



Fast

Industry leading performance. Latency under 200ms

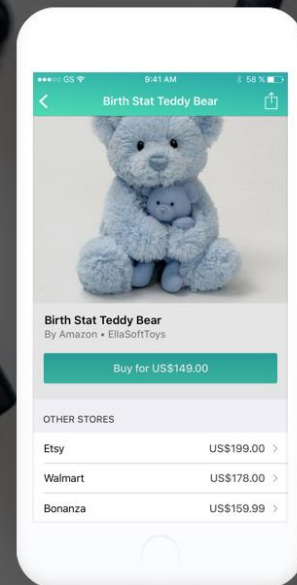
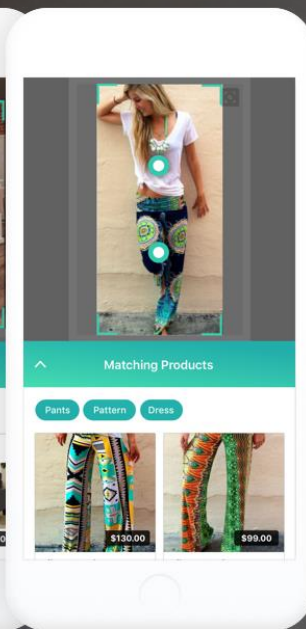
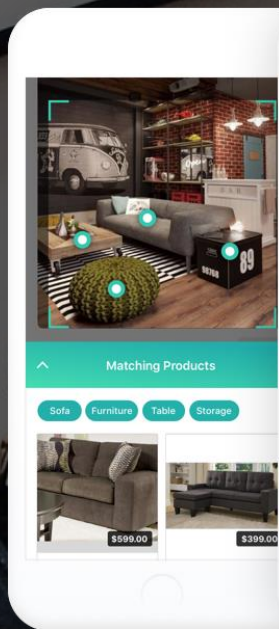
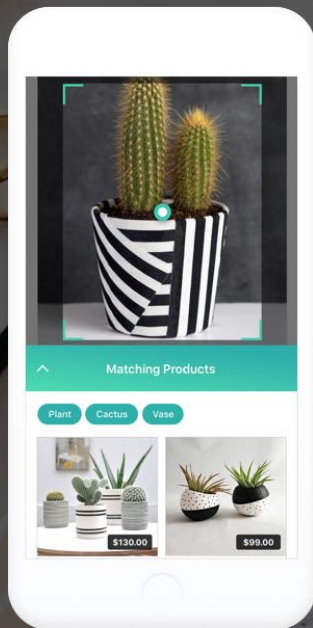
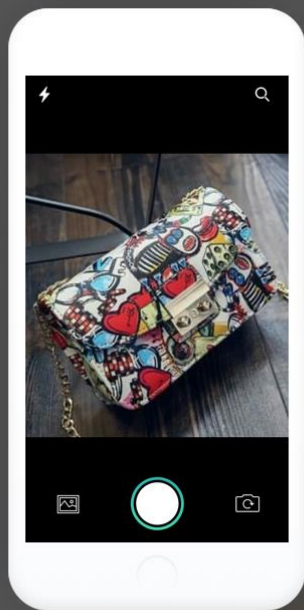


Accurate

Above 90% accuracy as per customer evaluations

Built-in failovers and professional services ensure high availability and support for troubleshooting

# EGAMI makes the camera shoppable



Download our Showcase app below :

<https://itunes.apple.com/us/app/egami/id1241860111?ls=1&mt=8>



Deep learning behind ViSenze's Cloud-based Visual Search and Recognition System

# Visual Search 1-2-3

Query image



Features  
(AlexNet)

$v[1:4096]$

Similarity + kNN  
(Cosine similarity)



$v_1$   
 $v_2$   
 $\dots$   
 $v_N$

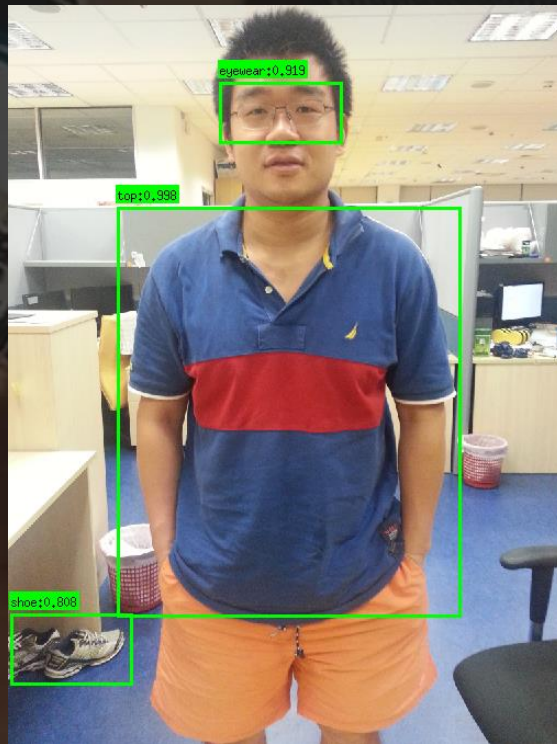
		
Similarity 78.752%	Similarity 78.267%	Similarity 77.827%
		
Similarity 77.473%	Similarity 77.207%	Similarity 76.673%



# Object Detection

What have we used?

- FFLD (sliding windows)
- Deformable Part Models
- Fast-RCNN
- Faster-RCNN
- SSD
- DSOD



# Visual Embedding

## Similarity:

- Exact match (the exact same item)
- Different variations of the same product
- Same category
- Similar category
- ...

Needs to combine multiple embeddings



Similarity 63.696%



Similarity 63.237%



Similarity 61.479%



Similarity 59.424%




Similarity 57.236%



Similarity 57.116%

## DL Problems ViSenze is working on

- CNN feature learning and indexing
- CNN Model compression
- Interpretation and visualization of CNN
- Robust fine-tuning and Incremental learning for various CV tasks
- Large-scale product category and attribute classification
- Automatic training/test data bias discovery

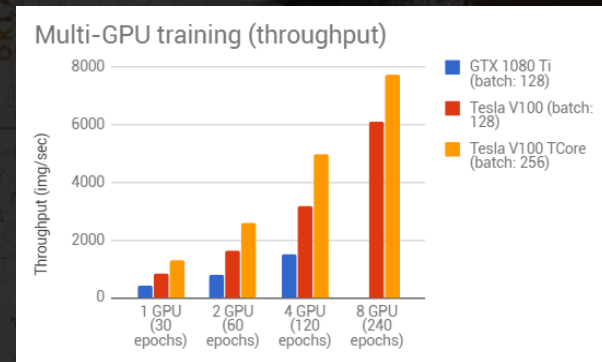
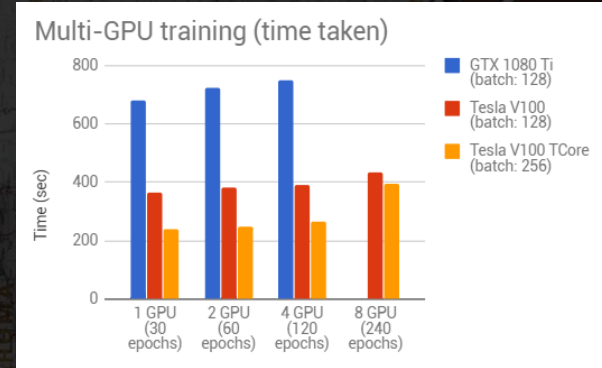
A top-down view of a desk with a map of the United States. Various items are scattered on the map, including a magnifying glass, a camera, a coffee cup, a pen, a notebook, a pair of shoes, and some papers. The text is overlaid on the map.

# ViSenze's Practices to Accelerate the R&D Efficiency using Deep Learning

# From Isolated GPU to Centralized Training Management



SYSTEM SPECIFICATIONS		
GPUs	8X Tesla V100	8X Tesla P100
TFLOPS (GPU FP16)	960	170
GPU Memory	128GB total system	
CPU	Dual 20-Core Intel Xeon ES-2698 v4 2.2 GHz	
NVIDIA CUDA* Cores	40,960	28,672
NVIDIA Tensor Cores (on V100 based systems)	5,120	N/A
Maximum Power Requirements	3,200 W	
System Memory	512 MB 2,133 MHz DDR4 LRDIMM	
Storage	4X 1.92 TB SSD RAID 0	
Network	Dual 10 GbE, Up to 4 IB EDR	
Software	Ubuntu Linux Host OS See Software Stack for Details	
System Weight	134 lbs	
System Dimensions	866 D x 444 W x 131 H (mm)	
Packing Dimensions	1,180 D x 730 W x 284 H (mm)	
Operating Temperature Range	10-35 °C	



- Pascal TitanX and 1080Ti, DGX-1 with 8 V100
- Move from single server management to centralized training management

A hand holding a smartphone displaying a street scene, with a semi-transparent dark overlay containing text.

## Visual Search and Recognition at ViSenze

Guangda Li  
[guangda@visenze.com](mailto:guangda@visenze.com)  
[www.visenze.com](http://www.visenze.com)