Machine Learning sur AWS

Julien Simon

Principal Technical Evangelist, AI and Machine Learning

@julsimon

A Long History of ML at Amazon

Personalized recommendations

Fulfillment automation and inventory management

Drones

Voice driven interactions

Inventing entirely new customer experiences

Our mission

Put Machine Learning in the hands
of every developer and data scientist

More ML is built on AWS than anywhere else

More ML is built on AWS than anywhere else

aws SUMMIT

Le marketing du futur ?
Le marketing individualisé

Olivier Marc, co-fondateur et CEO, Advalo

Le marketing de masse est polluant, intrusif, peu efficace

En moyenne,
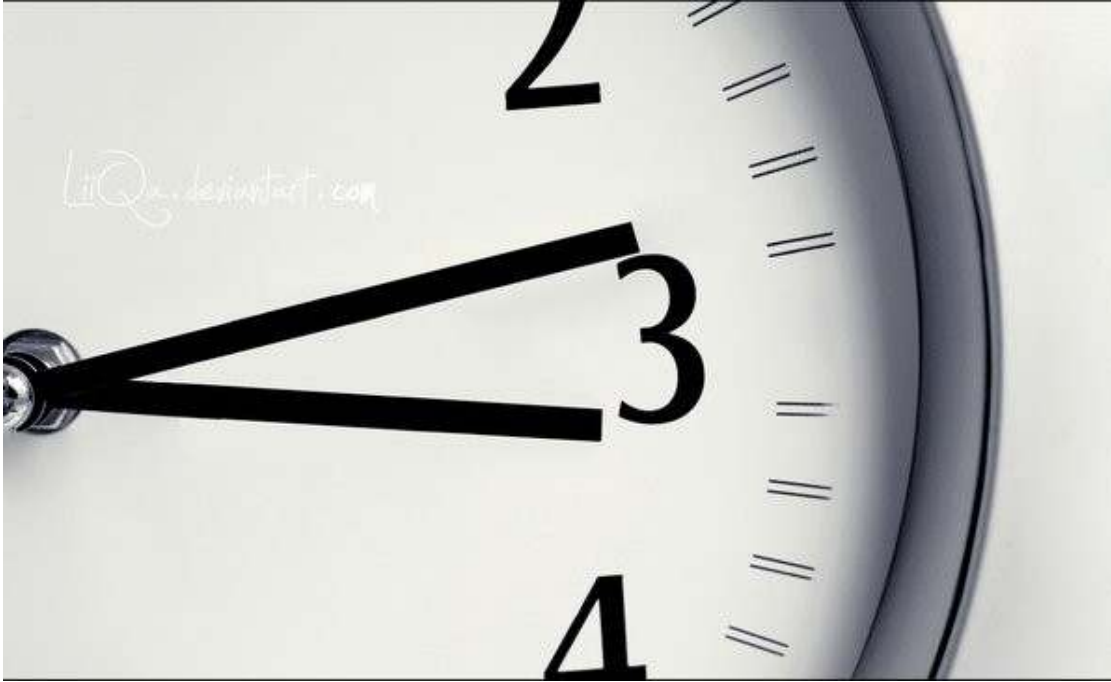
99%

des personnes ciblées ne réagissent pas

Organisées en silo, les entreprises sont incapables d'accompagner le consommateur au long de son parcours

En moyenne,

1%

des leads sont convertis

Le marketing de demain ? Un marketing individualisé qui intervient uniquement au bon moment

En moyenne,

70% des consommateurs ont pris leur décision d'achat AVANT d'aller en magasin

La combinaison data / IA est la seule solution pour répondre à ces nouveaux enjeux

advalo

- Création 2014, Paris/Rennes, 40+ personnes

- Editeur de la première plateforme de marketing individualisé

- Hyper croissance

- Série A de 5 millions d'euros fin 2017

CACHE CACHE.

Cuisines ixina

MORGAN
MORGAN DE TOI

BUT

SEAT

DEVRED 1902

ŠKODA

CÔTÉSUSHI
NIKKEI FROM PERU

Notre plate-forme technique

- Sur AWS depuis le premier jour
- 5 milliards de données traitées par jour
  Amazon EMR, Apache Spark
- 14 Teraoctets
  Amazon S3, Amazon ElastiCache, Amazon RDS
- 250 sources de données
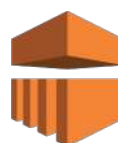- 150 modèles prédictifs en production

Amazon
EC2

Amazon
S3

Amazon
RDS

Amazon
ElastiCache

Amazon
EMR

# Use case #1 : SEAT

Enjeu : améliorer la conversion de *leads*

Challenge : exploiter l'ensemble des données pour scorer les intentionnistes actifs

Problème : le site génère beaucoup de bruit avec des amateurs de voiture qui n'ont pas particulièrement d'intentions d'achats

Solution : un modèle prédictif en temps réel identifie les intentionnistes
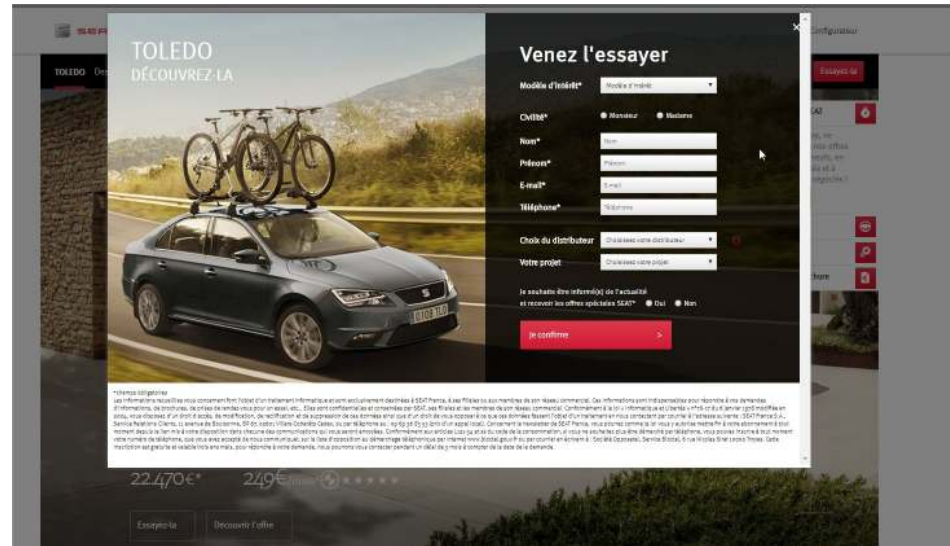
Une efficacité prometteuse
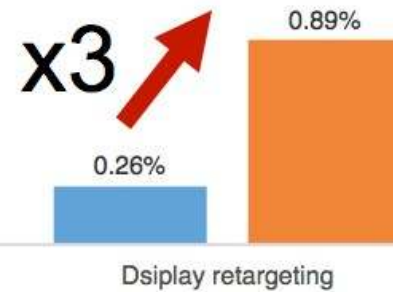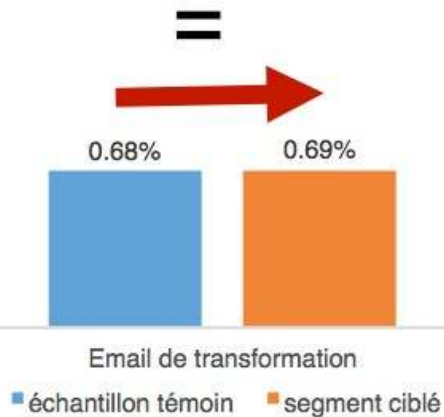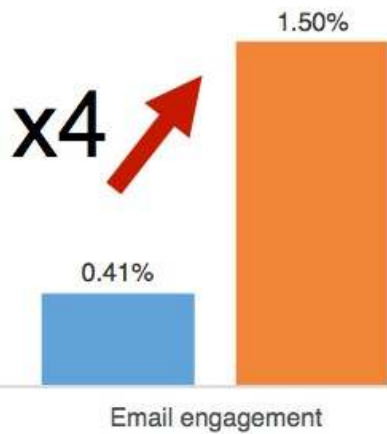
Use case #2 : Bessec

BESSEC
CHAUSSEUR DEPUIS 1862

Enjeu : générer du trafic en magasin

Challenge : individualiser et personnaliser la prise de parole dans le digital comme le ferait un *personal shopper*

Problème : les algorithmes de recommandation classiques ne fonctionnent pas lors des changements de collections

Solution: le Deep Learning au service de la détection de styles

Des résultats encourageants

5% des communications sont individualisées

60% des KPIs d'engagement

Prochaines étapes

Doublement de l'équipe

Ouverture d'un bureau aux USA début 2019

http://advalo.com

Merci !

AWS ML Stack

|	
Application Services	API-driven services: Vision, Language & Speech Services, Chatbots
Platform Services	Deploy machine learning models with high-performance machine learning algorithms, broad framework support, and one-click training, tuning, and inference.
Frameworks & Infrastructure	Develop sophisticated models with any framework, create managed, auto-scaling clusters of GPUs for large scale training, or run prediction on trained models.

Application Services

Amazon Rekognition

Deep Learning-based image analysis service

FREE TIER
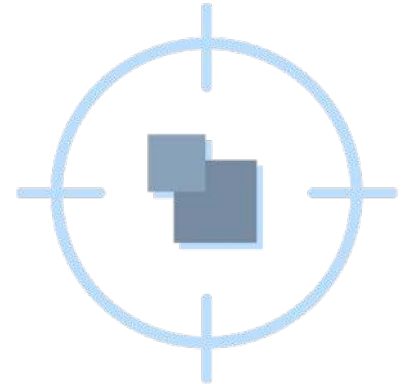
Object & Scene Detection

City	91.4%
Downtown	91.4%
Metropolis	91.4%
Urban	91.4%
Building	56.2%
High Rise	56.2%
Dock	55.6%
Pier	55.6%
Dawn	52.4%
Dusk	52.4%

Facial Analysis

looks like a face	99.8%
appears to be female	100%
age range	23 - 38 years old
smiling	99.4%
appears to be happy	93.2%
wearing eyeglasses	99.9%
wearing sunglasses	97.6%

Crowd Detection – up to 100 faces

Facial Search

Image Moderation

Results

Suggestive	83.5%
Female Swimwear Or Underwear	83.5%

Explicit Nudity
      Nudity
      Graphic Male Nudity
      Graphic Female Nudity
      Sexual Activity
      Partial Nudity
Suggestive
      Female Swimwear or Underwear
      Male Swimwear or Underwear
      Revealing Clothes

Celebrity Recognition

Text in Image

Amazon Rekognition Video

Deep Learning-based video analysis service
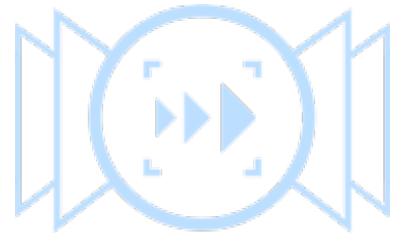
FREE TIER

Video Analysis
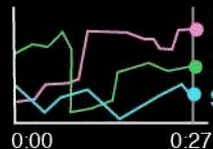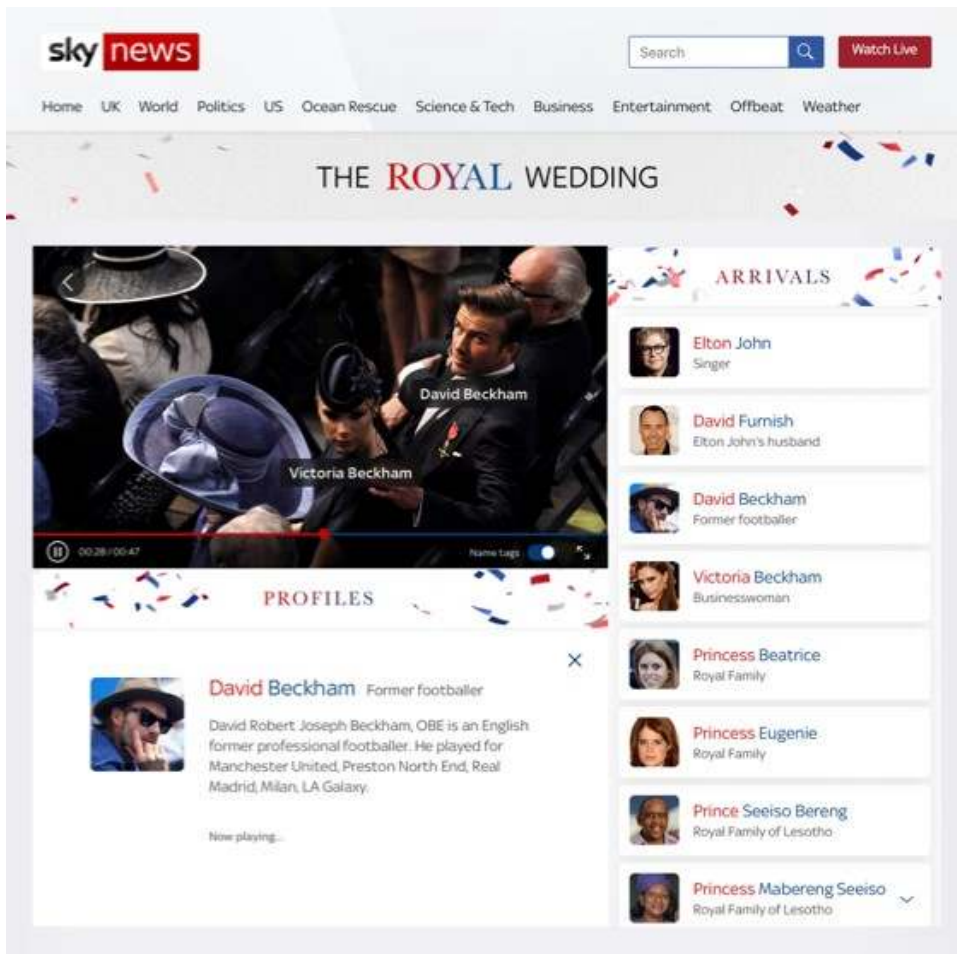
Sky News will use Amazon Rekognition to perform real-time identification of guests as they enter St. George's Chapel

Amazon Polly

Deep Learning-based text-to-speech service

FREE TIER

Amazon Polly: Text In, Life-like Speech Out

"Salut, je m'appelle Léa. Je suis la nouvelle voix française de Polly."

Amazon Polly
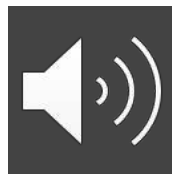
A Focus On Voice Quality & Pronunciation

Support for Speech Synthesis Markup Language (SSML) Version 1.0
https://www.w3.org/TR/speech-synthesis

```
<speak xml:lang="en-US">

    The price of this book is <prosody rate="60%">45€</prosody>

</speak>
```

Breathing

https://aws.amazon.com/blogs/machine-learning/amazon-polly-releases-new-ssml-breath-feature/

<amazon:auto-breaths>
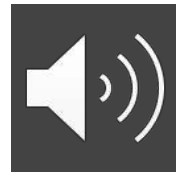
Amazon Translate

Neural Machine Translation Service

Now generally available

Amazon Translate

Natural and fluent language translation

Real-time
translation

Batch analysis

Automatic language
recognition

Low cost

Languages

English to/from:

- Spanish
- Portuguese
- German
- French
- Arabic
- Simplified Chinese

- Japanese
- Russian
- Italian
- Traditional Chinese
- Turkish
- Czech

Amazon Transcribe

Automatic speech recognition service

Now generally available

Amazon Transcribe

Automatic conversion of speech into accurate, grammatically correct text

Multiple languages

Intelligent punctuation and formatting

Timestamp generation
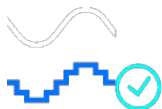
Support for telephony audio

Recognize multiple speakers

Custom vocabulary

Amazon Comprehend

Natural Language Processing

FREE TIER

Fully managed natural language processing

Discover valuable insights from text

Entities

Key Phrases

Amazon
Comprehend

Language

Sentiment

API explorer

Paste the text that you would like to analyze with natural language processing.

Clear text

The accompanying shots from the `Newcastle Herald` 's award-winning team of photographers are worth `more than 1000 words` apiece. But that's the maximum number of words you will require to have a shot at winning the `Herald` 's `sixth annual short st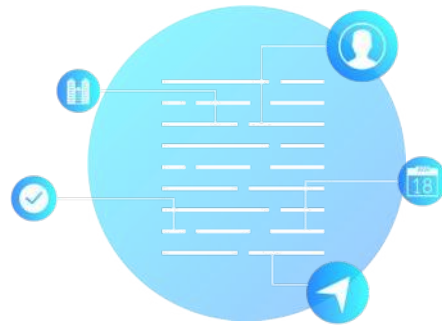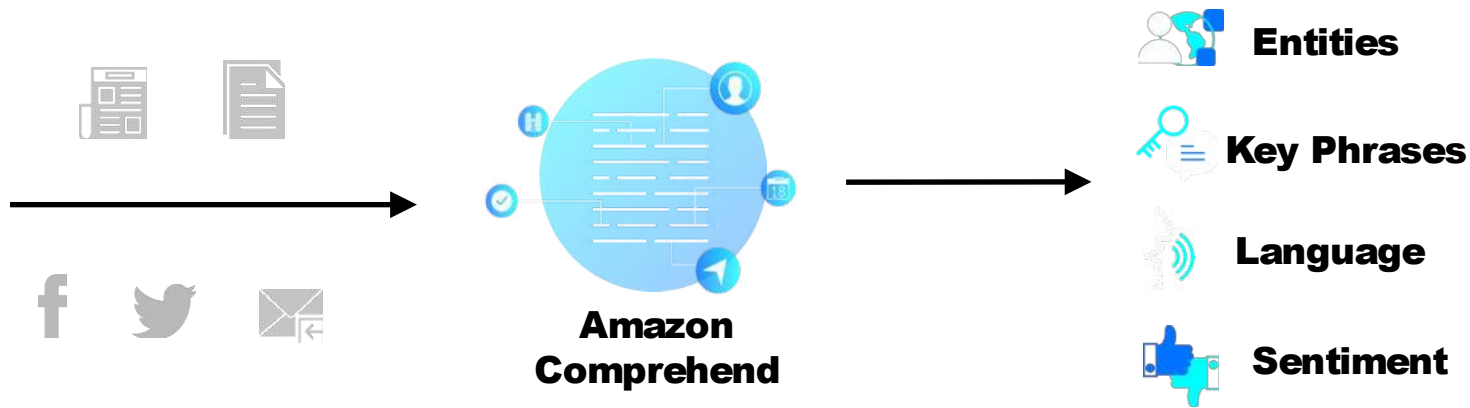ory competition` . It's time to let your inner author off the leash and allow your imagination to run wild. Stories can be of any genre but must be your own work. The plot must draw inspiration in some way from `one` of these `four images` . Herald photographers `Simone De Peak` , `Marina Neil` , `Jonathan Carroll` and `Max Mason-Hubers` captured the evocative pictures. Stories submitted for the competition must be between `900` and `1000 words` . Judges include `Newcastle Writers Festival` director `Rosemarie Milsom` , `Herald` deputy editor `Matt Carr` , `Hunter Writers Centre` director `Karen Crofts` and president `Megan Buxton` , and `Fairfax Media` 's `Newcastle-Hunter` group managing editor `Chad Watson` .

4

916 of 1000 characters used

Language   English

Analyze

Entity

This API returns the named entities ("Person", "Organization", "Locations", etc.) within the text you analyzed.

List   Tiles   JSON   Filter   Show all categories

Entity	Category	Count	Confidence
Newcastle Herald	● Organization	1	0.89
more than 1000 words	● Quantity	1	0.91
Herald	● Organization	2	0.46
sixth annual short story competition	● Quantity	1	0.79
one	● Quantity	1	0.4
four images	● Quantity	1	0.96
Simone De Peak	● Person	1	0.99
Marina Neil	● Person	1	0.99+
Jonathan Carroll	● Person	1	0.99+
Max Mason-Hubers	● Person	1	0.99
900	● Quantity	1	0.99

Show all

Topic modeling

LIBRARY OF
NEWS ARTICLES *

**Amazon
Comprehend**

STORM

STOCK
MARKET

AUSTRALIA   HEALTH

WORLD
SERIES

WASHINGTON

CRISIS

MACHINE
LEARNING

Platform Services

ML is still too complicated for everyday developers

Collect and prepare training data

Choose and optimize your ML algorithm

Set up and manage environments for training

Train and tune model (trial and error)

Deploy model in production

Scale and manage the production environment

Amazon SageMaker

Build	Train	Deploy
Pre-built notebooks for common problems Built-in, high-performance algorithms	One-click training Hyperparameter optimization	One-click deployment Fully managed hosting with auto-scaling

FREE TIER

Frameworks & Infrastructure

Amazon EC2 C5 instances

C5: Next Generation Compute-Optimized Instances with Intel® Xeon® Scalable Processor

AWS Compute optimized instances support the new Intel® AVX-512 advanced instruction set, enabling you to more efficiently run vector processing workloads with single and double floating point precision, such as AI/machine learning or video processing.

25% improvement in price/performance over C4

C4

36 vCPUs

(intel) "Haswell"

4 Gbps to EBS

60 GiB memory

C5

72 vCPUs

2X vCPUs

(intel) "Skylake"

12 Gbps to EBS

3X throughput

144 GiB memory

2.4X memory

AVX 512

2X performance

(intel) XEON PLATINUM inside™
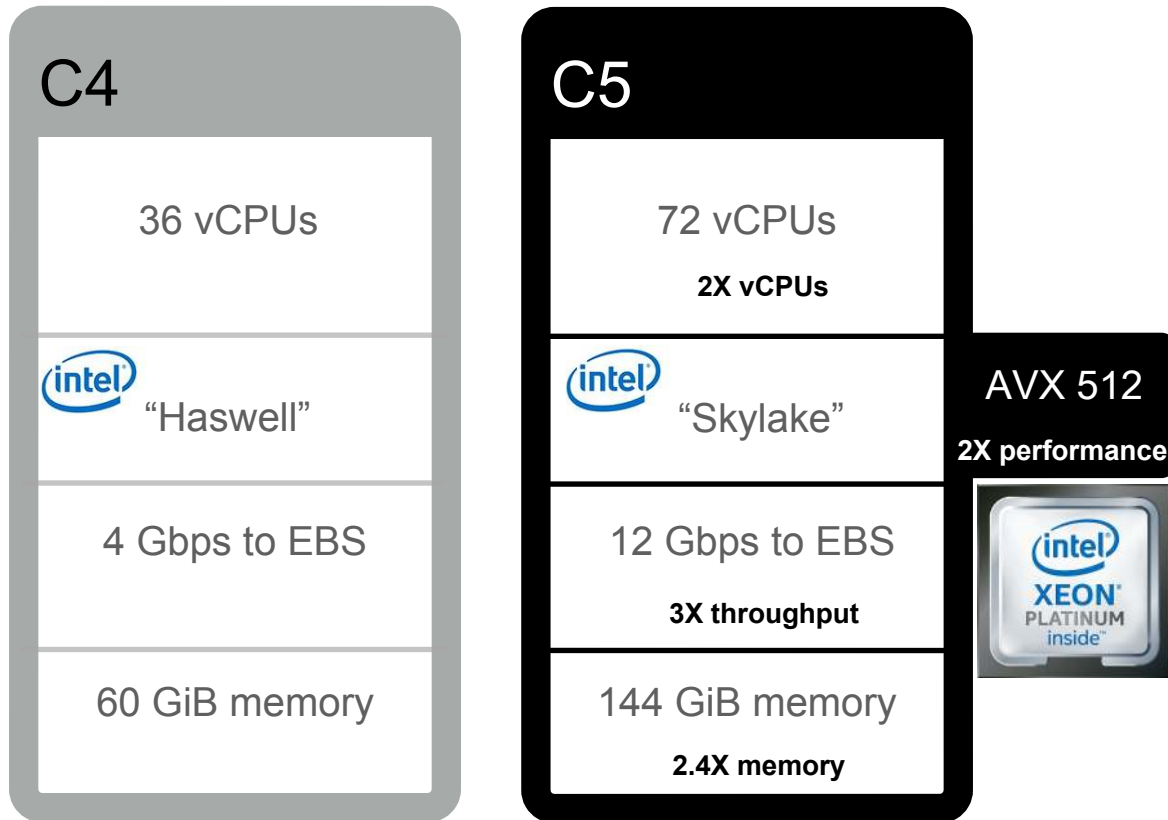
Amazon EC2 P3 Instances

The fastest, most powerful GPU instances in the cloud

- P3.2xlarge, P3.8xlarge, P3.16xlarge

- Up to eight NVIDIA Tesla V100 GPUs in a single instance
  - 40,960 CUDA cores, 5120 Tensor cores
  - 128GB of GPU memory

- 1 PetaFLOPs of computational performance *– 14x better than P2*

- 300 GB/s GPU-to-GPU communication (NVLink) *– 9x better than P2*

AWS Deep Learning AMIs

Preconfigured environments to quickly build Deep Learning applications

Conda AMI
For developers who want pre-installed pip packages of DL frameworks in separate virtual environments.

Base AMI
For developers who want a clean slate to set up private DL engine repositories or custom builds of DL engines.

AMI with source code
For developers who want preinstalled DL frameworks and their source code in a shared Python environment.
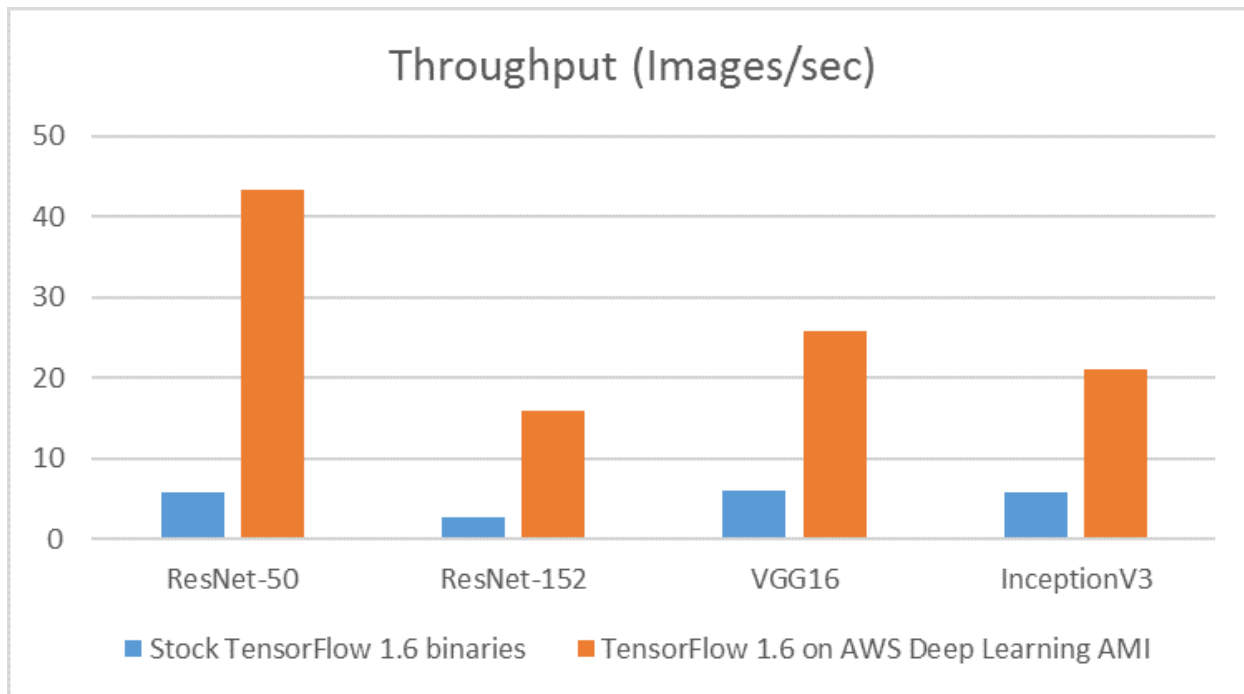
https://aws.amazon.com/machine-learning/amis/

Faster TensorFlow training on C5

Throughput (Images/sec)

Can we do more to put ML in the hands of all developers (literally)?

AWS DeepLens

The world's first Deep Learning-enabled video camera for developers

HD video camera

Custom-designed Deep Learning engine

intel®

Micro-SD
Mini-HDMI
USB
USB
Reset
Audio out
Power
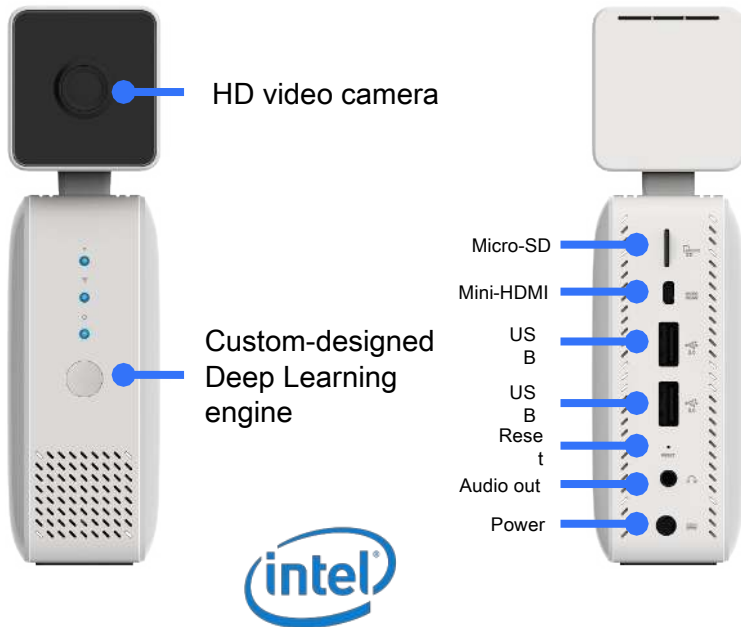
HD video camera with on-board compute optimized for Deep Learning

Integrates with Amazon SageMaker and AWS Lambda

10 MIN

From unboxing to prediction in <10 minutes

Tutorials, examples, demos, and pre-built models

Object detection with AWS DeepLens

Democratization of ML

APPLICATION SERVICES

Amazon Rekognition

Amazon Rekognition Video

Amazon Polly

Amazon Transcribe

Amazon Lex

Amazon Translate

Amazon Comprehend

PLATFORM SERVICES

Amazon SageMaker

AWS DeepLens

FRAMEWORKS AND INTERFACES

Deep Learning AMI

| Apache MXNet | Caffe2 | CNTK | PyTorch | TensorFlow | Theano | Torch | Keras | Gluon |

Thank you!

https://aws.amazon.com/machine-learning
https://aws.amazon.com/blogs/ai

https://medium.com/@julsimon
https://youtube.com/juliensimonfr

Julien Simon

Principal Technical Evangelist, AI and Machine Learning

@ julsimon