AWS re:Invent

AIM405

Optimize deep learning models for edge deployments with AWS DeepLens

Phu Nguyen
Product Manager, AI Devices
Amazon Web Services

Nathaniel Slater
Sr. Manager, ML Solutions Lab
Amazon Web Services

AWS DeepLens
is not a
video camera

It's the
world's first
deep learning
enabled
developer kit

intel
inside™

AWS DeepLens

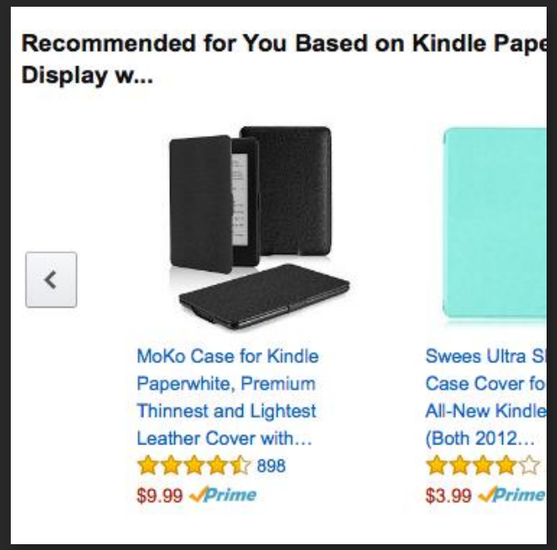
The world’s first deep learning–enabled video camera for developers

- Purpose-built for ML skills development
- Fully programmable and customizable
- Build custom Amazon SageMaker models
- 10 minutes to your first deep-learning project

Agenda

- Machine learning at AWS
- Lab 1: Label data with Amazon SageMaker Ground Truth
- Lab 2: Train a custom model with Amazon SageMaker
- Lab 3: Deploy custom model to AWS DeepLens
- Lab 4: Optimize model for edge devices with Amazon SageMaker Neo

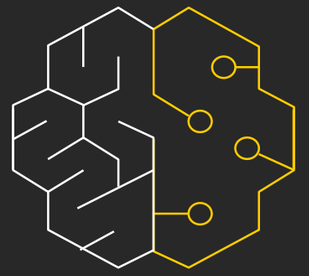
Amazon’s ML innovation



Our mission at AWS

Put machine learning in the
hands of every developer

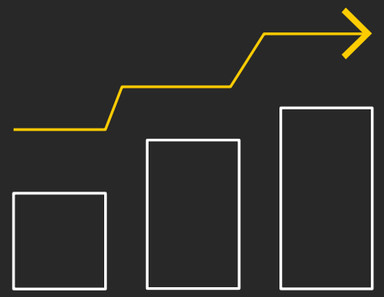
Why AWS for ML?



Broadest and deepest set of AI and ML services

200 new features and services launched in the last year alone

Unmatched flexibility



Accelerate your adoption of ML with Amazon SageMaker

70% cost reduction in data labeling

10x faster performance

75% lower inference cost

Built on the most comprehensive cloud platform

AWS named as a leader in Gartner's Infrastructure as a Service (IaaS) Magic Quadrant for the 9th consecutive year

More machine learning happens on AWS than anywhere else

More than 10,000 customers | 2x the customer references | 85% of TensorFlow projects in the cloud happen on AWS

The AWS ML stack

Broadest and deepest set of capabilities

AI services

Vision			Speech		Language		Chatbots	Forecasting	Recommendations
Amazon Rekognition Image	Amazon Rekognition Video	Amazon Textract	Amazon Polly	Amazon Transcribe	Amazon Translate	Amazon Comprehend and Amazon Comprehend Medical	Amazon Lex	Amazon Forecast	Amazon Personalize

ML services

Amazon SageMaker

Ground Truth	Notebooks	Algorithms + AWS Marketplace	Reinforcement learning	Training	Optimization	Deployment	Hosting

ML frameworks + infrastructure

Frameworks	Interfaces
TensorFlow, mxnet, PYTORCH	GLUON, Keras

Infrastructure

Amazon EC2 P3 and P3DN
Amazon EC2 G4
Amazon EC2 C5
FPGAs
DL containers and AMIs
Amazon Elastic Container Service (Amazon ECS)
Amazon Elastic Kubernetes Service (Amazon EKS)
AWS IoT Greengrass
Amazon Elastic Inference
AWS Inferentia

Enabling the next machine learning developers

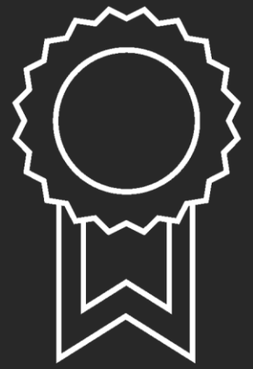
ML learning devices

AWS DeepLens
(deep learning)

AWS DeepRacer
(reinforcement
learning)

AWS DeepComposer
(generative AI)

Training and certification



AWS ML training and
certification

UDACITY

coursera

edX

Partnerships with
MOOCs

Let’s hear from one of our developers

Workshop labs

AWS re:Invent

aws

What you will build today

In search of food, polar bears are wandering further into human towns every year. Encounters between humans and bears can be deadly for both.

Our mission is to create an early bear detection system that sends an SMS message to a nearby ranger as soon as a bear is spotted.

Bear: 99.9%, Cuteness*: 70.2%

*Cuteness not a reliable prediction by machine learning

Photo by Brain McMahon on Unsplash

Lab 1: Label data with Amazon SageMaker Ground Truth

- Instructions
  - Go to Amazon SageMaker > deeplens notebook
  - Go to AIM405 > Lab 1 > deeplens-400-lab1-gt.ipynb
- Goals
  - Initiate a private labeling job for 10 photos with Amazon SageMaker Ground Truth
  - We will provide a bigger dataset for training/validation
- Time
  - 10 minutes

Lab 2: Train GluonCV model

- Instructions
  - Go to Amazon SageMaker > deeplens notebook
  - Go to AIM405 > Lab 1 > lab2-image-classification.ipynb
- Goals
  - Use a pretrained GluonCV from the Gluon model zoo
  - Use Amazon SageMaker to do transfer learning to train a classification model to detect if an image has a bear in it or not
  - Compile model for Amazon SageMaker Neo
- Time
  - 30 minutes

Considerations for edge devices

- Spotty internet connectivity
- Extreme environments
- Computer power/ performance tradeoff

Photo by Todd Torabi on Unsplash

Lab 3: Deploy to AWS DeepLens

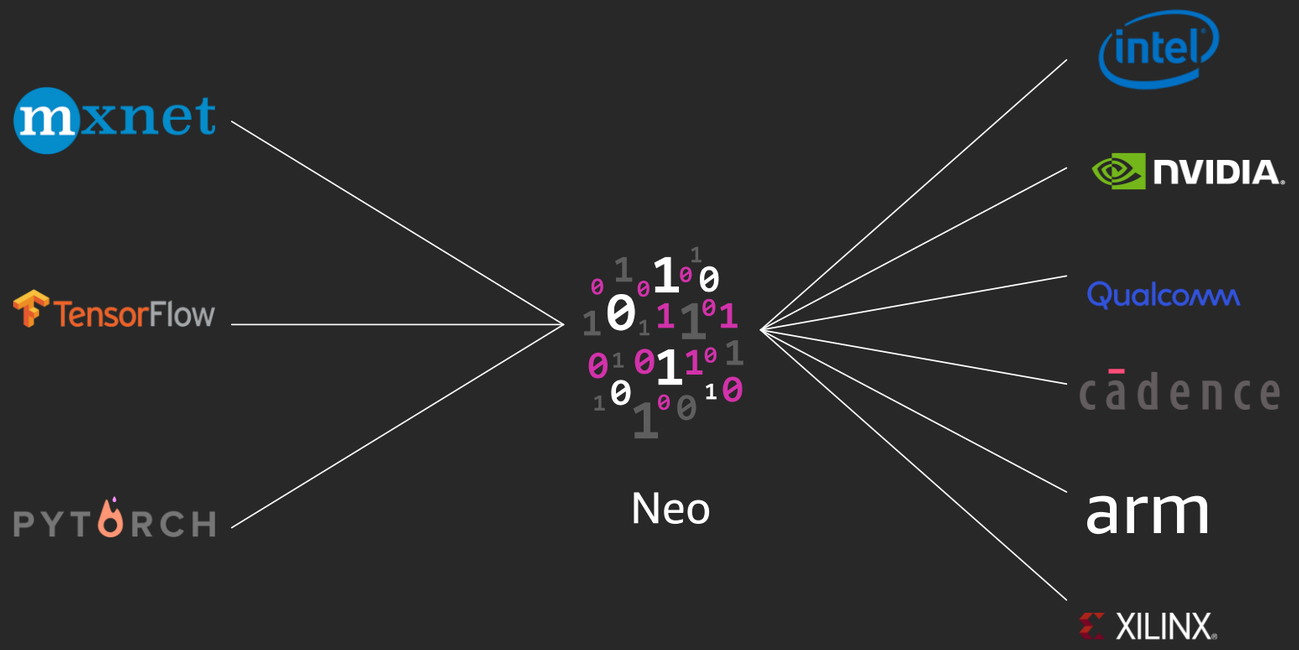
- Instructions
  - https://github.com/aws-samples/aws-deeplens-reinvent-2019-workshops
  - Go to AIM405 > Lab 3
- Goals
  - Write a Lambda function to do inference on AWS DeepLens
  - Package the model artifacts and deploy to AWS DeepLens
  - See inference results in IoT console
- Time
  - 20 minutes

Lab 4: Amazon SageMaker Neo

- Instructions
  - https://github.com/aws-samples/aws-deeplens-reinvent-2019-workshops
  - Go to AIM405 > Lab 4
- Goals
  - Compile a model optimized for edge deployment with Amazon SageMaker Neo
  - Deploy optimized model to AWS DeepLens
- Time
  - 15 minutes

Amazon SageMaker Neo

Train once, run anywhere



Open-source device runtime and compiler,
1/10th the size of original frameworks

Conclusion

What you've built

- Labeled a dataset with Amazon SageMaker Ground Truth
- Trained a classification model using transfer learning with Amazon SageMaker and GluonCV
- Deployed custom model to AWS DeepLens

What else can you build?

ReadToMe
Created by Alex Schultz

ReadToMe is a deep-learning-enabled application can read books to kids. In this case, reading *Green Eggs and Ham* by Dr. Seuss.

Dee
Created by Matthew Clark

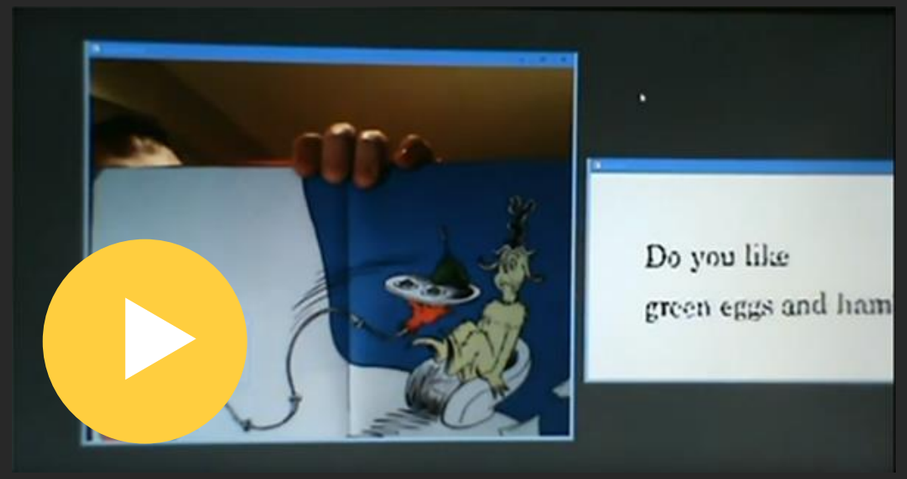
Dee is a fun AWS DeepLens interactive device for children. The device asks children to answer questions by showing a picture of the answer.

SafeHaven
Created by Nathan Stone and Peter McLean

SafeHaven uses Alexa and AWS DeepLens to bring peace of mind to vulnerable people and their families

Access the GitHub repos and more use cases: https://aws.amazon.com/deeplens/community-projects

Spotlight project: ReadToMe



ReadToMe
Created by Alex Schultz

ReadToMe is a deep-learning-enabled application that can read books to kids. In this case, reading *Green Eggs and Ham* by Dr. Seuss.

Solving problems with AWS DeepLens

"I had a problem that kept occurring, so I thought I'm going to try this. AWS DeepLens got me started with ML quickly."

Ben Hamm, Sr. Product Manager—Technical, AWS

Get your AWS DeepLens today

Special offer for DeepLens workshop participants

~~$249~~

$99

Available on Amazon.com

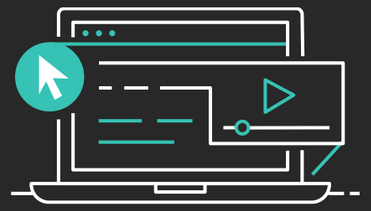
https://aws.amazon.com/deeplens/

Questions?

Learn ML with AWS Training and Certification

The same training that our own developers use, now available on demand

Role-based ML learning paths for developers, data scientists, data platform engineers, and business decision makers



70+ free digital ML courses from AWS experts let you learn from real-world challenges tackled at AWS

Validate expertise with the
AWS Certified Machine Learning - Specialty exam

Visit https://aws.training/machinelearning

aws training and certification

Thank you!

Phu Nguyen

phu@amazon.com

Nathaniel Slater

slatern@amazon.com

AWS re:Invent

aws

Please complete the session survey in the mobile app.