

# Дослідження використання безсерверних технологій для навчання нейронних мереж

Виконав: Воловик І. М.

Керівник: Науменко Т. О.

# Актуальність досліджень

- Висока вартість інфраструктури для навчання нейронної мережі
- Високий поріг входу для Data Science інженерів
- Складність розміщення нейронної мережі у великих проектах

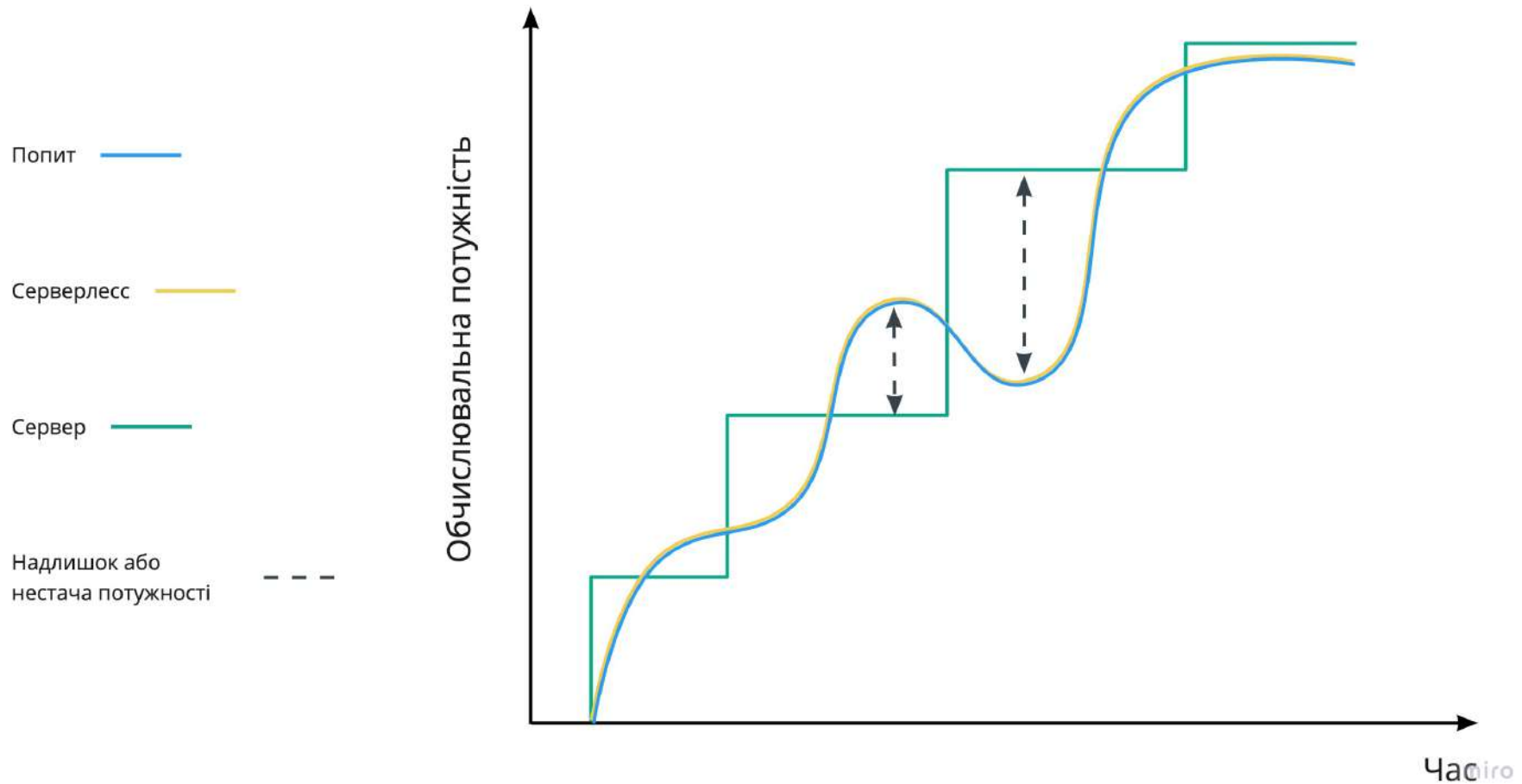
# Мета роботи

Дослідити використання безсерверних технологій, зокрема FaaS з нейронними мережами у великих проектах

# Постановка задачі

- Аналіз
  - Зниження коштів
  - Спрощення навчання
- Розробка
  - Безсерверний аналіз відеопотоку
  - Розробка аналогу AWS Recognition
  - Детекція медичних масок
  - Безсерверний аналіз повідомлень із соціальних мереж

# Зниження коштів за допомогою AWS Lambda



# Зниження коштів за допомогою AWS Spot

Різні схеми оплати під різні задачі:

- On-demand
- Reserved
- Spot



# Спрощення навчання за допомогою Spotty

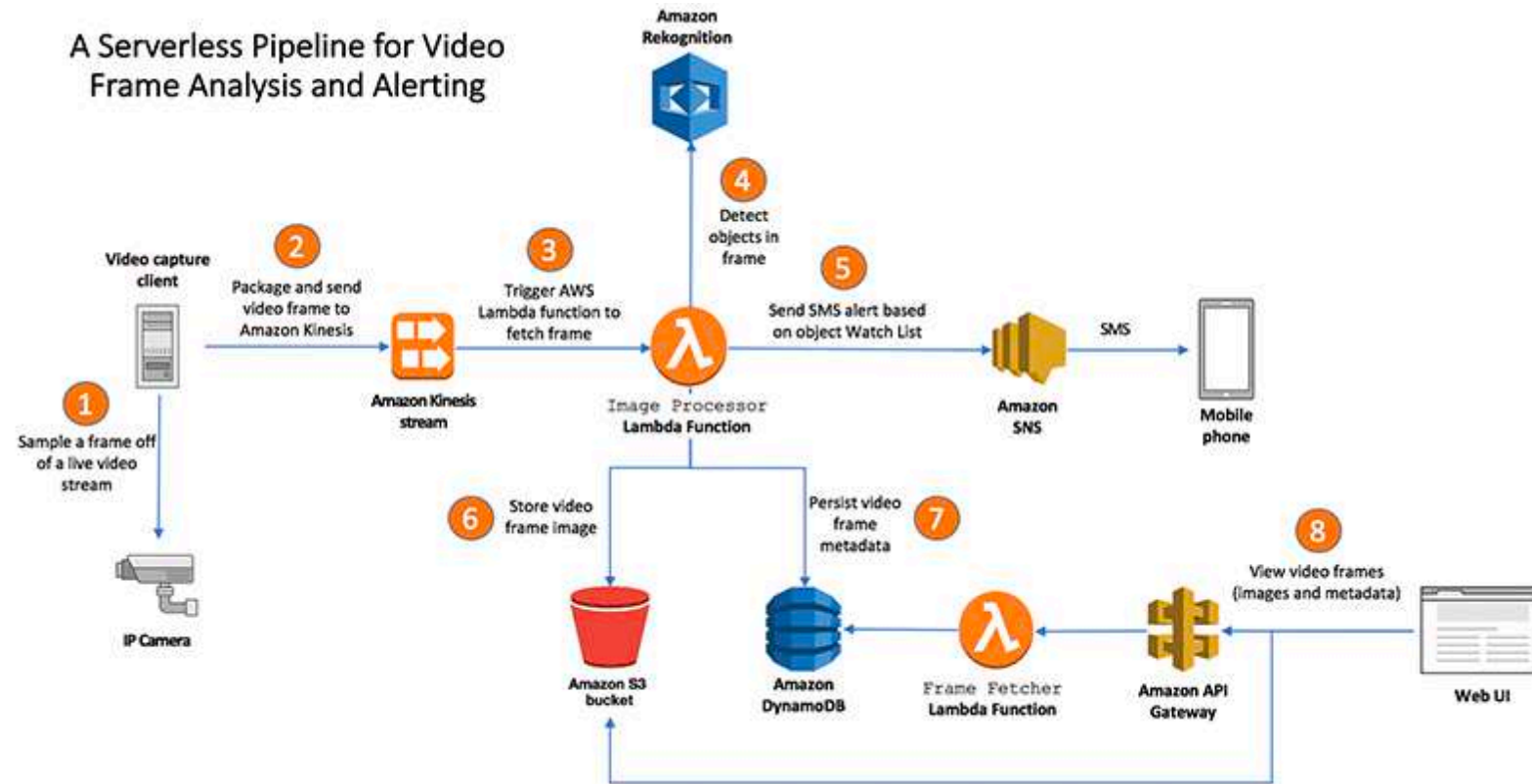
- Налаштування за 5 хвилин через уатл файл
- Потребує мінімальних знань клауд провайдеру
- До 10% зміни налаштувань при переході до іншого провайдеру

The instance was successfully started.

Instance State	running
Instance Type	p2.xlarge
Availability Zone	us-east-2a
Public IP Address	3.16.79.217
Launch Time	2019-04-13 19:21:15
Purchasing Option	Spot Instance
Spot Instance Price	\$0.2870

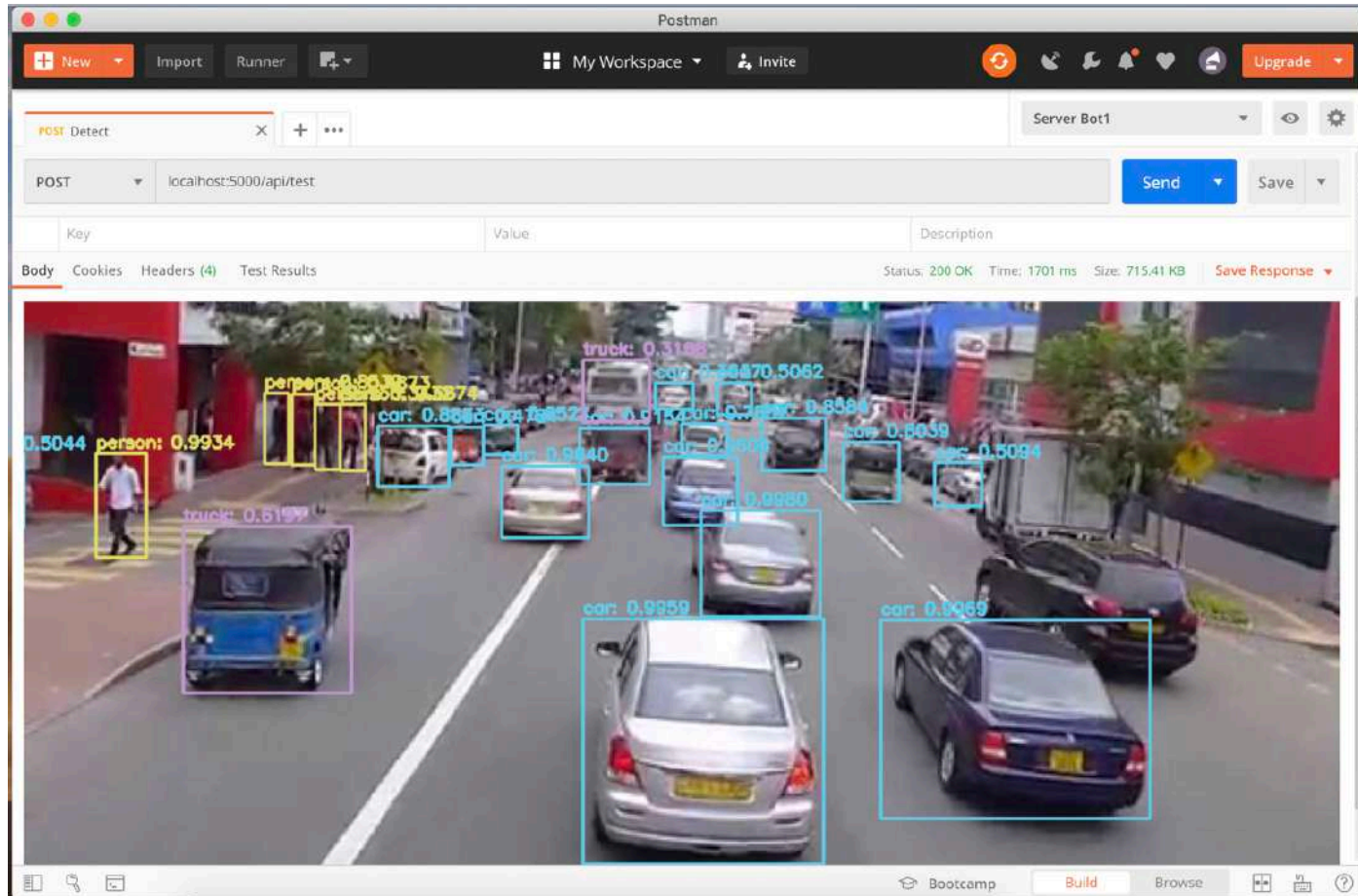
Use the "spotty ssh" command to connect to the Docker container.

# Безсерверний аналіз відеопотоку з детекцією

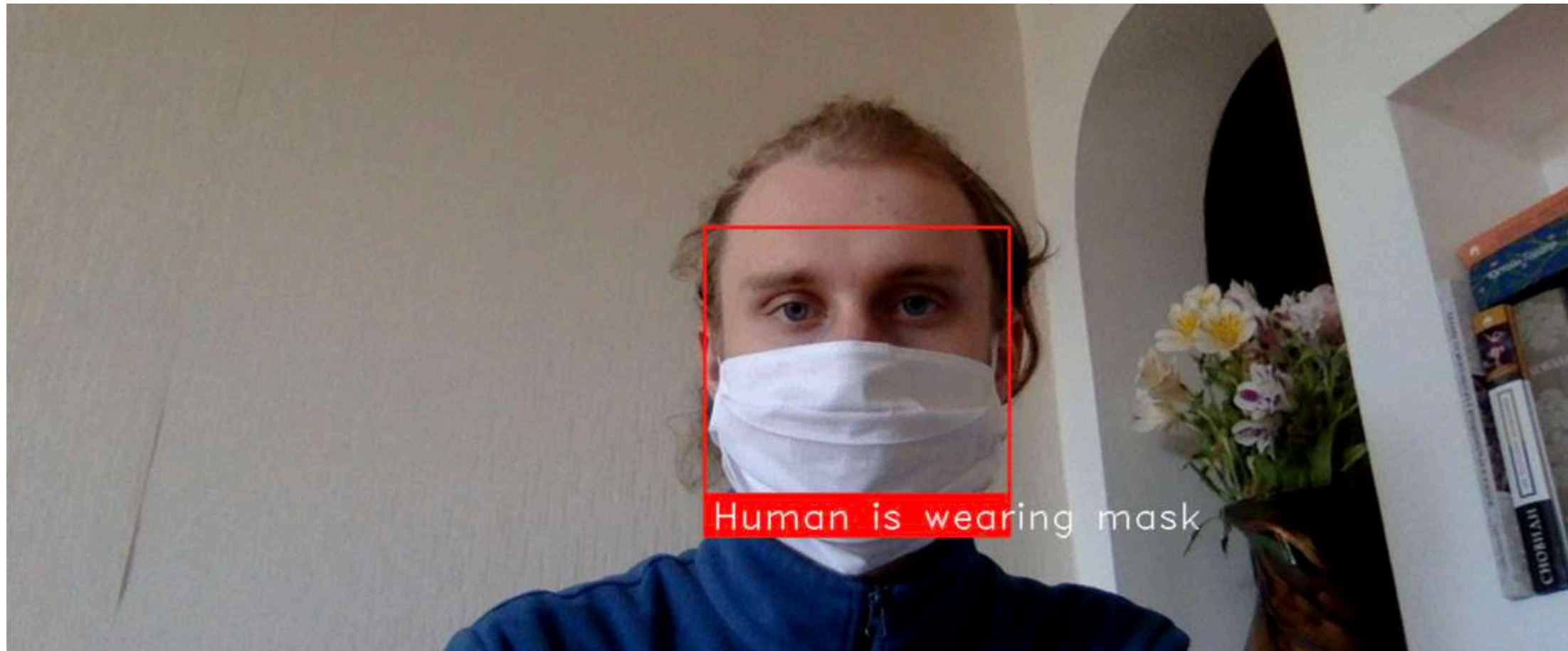




# Розробка аналогу AWS Recognition



# Детекція медичних масок





# Об'єднання обох підходів

- Винесення нейронної мережі на окрему VM при постійному навантаженні на неї
- Перетворення компонентів які мають змінне навантаження на безсерверні компоненти
- Утримання своєї бази даних для менш важливих даних, оскільки DBaaS з кросрегіональною реплікацією дорожчі за їх серверні аналоги

# Висновки

- Навчання нейронної мережі потребують значних коштів, які можна знизити за допомогою серверлесс рішень
- Нейронна мережа у великих проектах потребує допоміжної інфраструктури, а також людей і ресурсів для її підтримування та розробки, що можна спростити використавши FaaS та інші серверлесс компоненти
- На прикладі розроблених додатків було виявлено:
  - Повністю безсерверне рішення не завжди є найефективнішим, найкраще показують себе моделі які поєднують серверний та безсерверний підходи
  - Частина побудованого серверлесс рішення можна швидко замінити покращивши рішення або отримавши систему для вирішення іншої, сумісної задачі