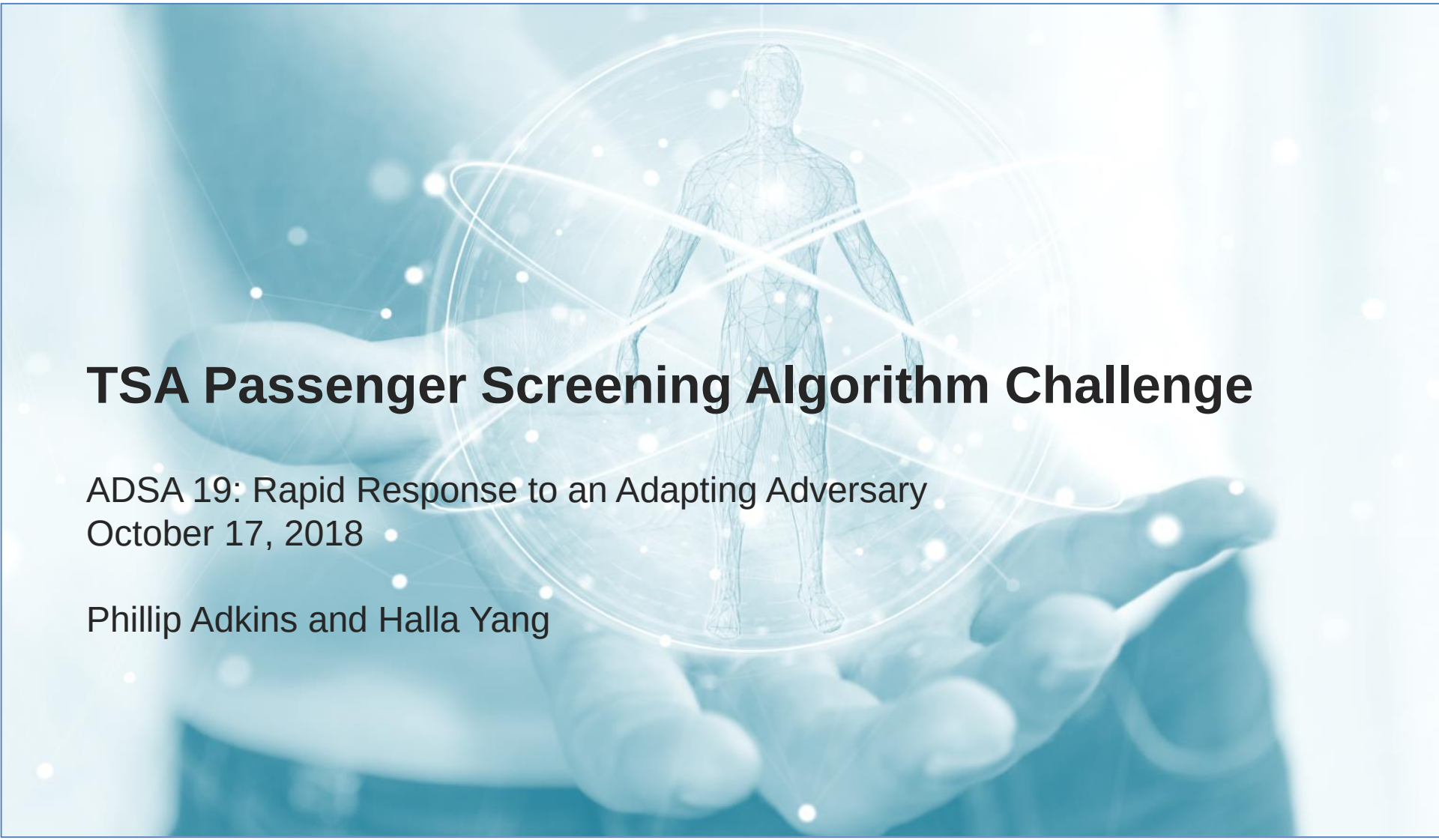


TSA Passenger Screening Algorithm Challenge

ADSA 19: Rapid Response to an Adapting Adversary
October 17, 2018

Phillip Adkins and Halla Yang

kaggle

PASSENGER SCREENING ALGORITHM CHALLENGE

Background

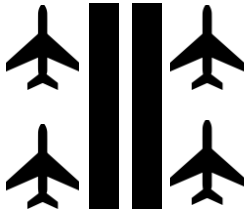
Phillip Adkins
Chicago, IL

- B.S. **Physics** (University of Southern California)
- 6+ years data science at **Citadel**, **Banjo, Gamut**, …
- Prior **Kaggle prize wins**:
  - **2nd Place**, Epilepsy Seizure Prediction Challenge

Halla Yang
Chicago, IL

- B.A. **Physics** (Harvard)
- Ph. D. in **Business Economics** (Harvard)
- 10+ years at **Citadel, McKinsey, Goldman Sachs, …**
- Prior **Kaggle prize wins**:
  - **1st Place**, Pfizer Prescription Volume Prediction
  - **2nd Place**, Recruit Coupon Purchase Prediction

Motivation



- **Opportunity** to help improve airport security.
- Remembrance of **9/11.**

OAK RIDGE National Laboratory

- **Family tradition**: parents both conducted research at ORNL, other labs.
- Both worked on security related research at some point in their careers.

- Had previously obtained 4 of 5 “gold medals” required on Kaggle to become ‘**grandmaster**,’ an achievement rank attained by only 100 competitors to date.
- This competition was the 5th gold medal.

kaggle

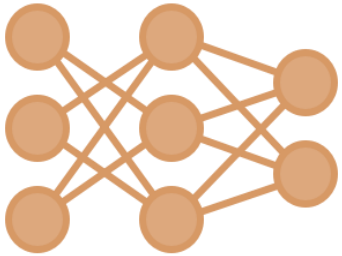
PASSENGER SCREENING ALGORITHM CHALLENGE

Approach

Step 1. Image Processing

- **Random forest** classifier applied to find **hands, heads, legs**, etc.
- Threats are usually only visible in certain angles, so need crops from multiple viewpoints.
- This step reduces dimensionality of the original problem and makes further training more likely to succeed.

Step 2. Feature Extraction

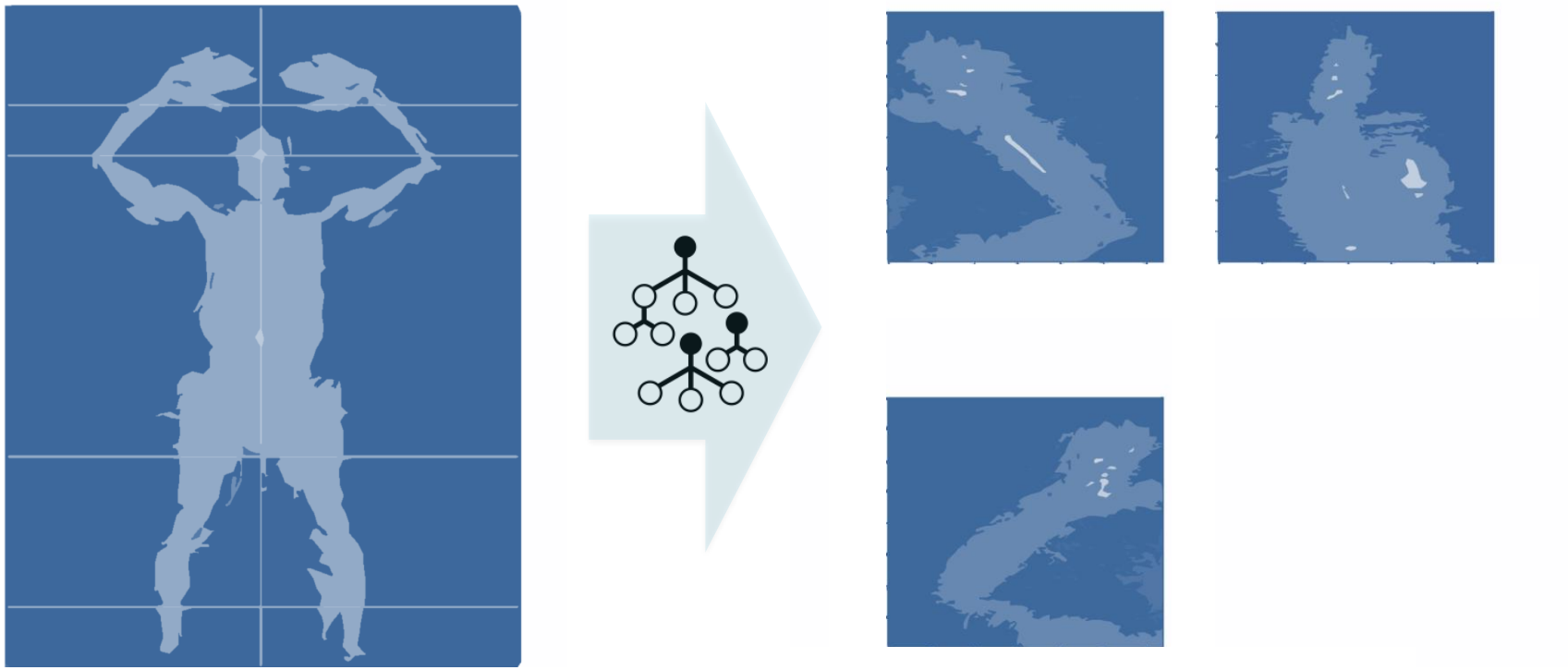


- Each crop is fed through a traditional conv/pooling architecture.
  - — Dimensionality is reduced at each step.
- Features extracted from final pooling layer fed through dense layers.
  - — Dropout is used to avoid over-fitting when training.
- One-hot encoding of zone ID is merged into dense layers.

Step 3. Prediction

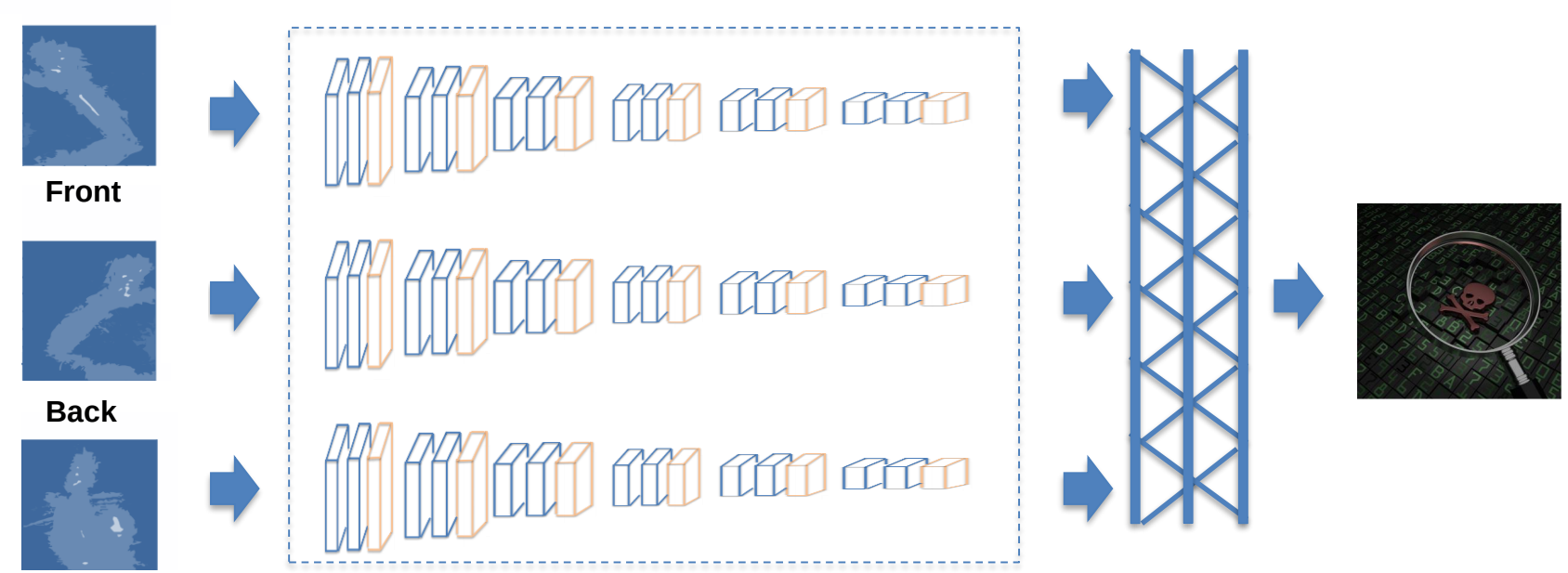
- Predictions from separately trained models are averaged together.
- Final score: **6th place** with log loss 0.055 ($100K prize).
- Solution **likely to generalize out of sample.**
  - Did **not** rely on **contest-specific** assumptions.

Step 1: Image Processing



- We **trained a random forest classifier** to automatically identify body segments, e.g. upper right arm, lower left hip, that correspond to TSA-identified zones of interest.
- The classifier used **symmetry** and searched for **centers of mass**, producing the likely horizontal and vertical coordinates of the midpoint of each zone (e.g. lower left arm in this example) in all angles.
- Because the crops **were likely to be imperfect**, we applied translations, stretches, and other perturbations when training our networks to **increase robustness**.

Steps 2 - 3: Feature Extraction and Prediction



- **Convolutional neural network** is trained from scratch – no use of pre-trained models.
- Weights of networks applied to the different views are tied together, since the identification of threats should be **invariant to angle.**
- **Training** time ranged from 24 hours for the 3-input network to 48 hours for the 7-input network on Amazon Web Services GPU instances.
- Predictions of 3-view, 6-view and 7-view angles blended together.
- Prediction on new images is **fast,** implementable in **real time** with **minimal latency.**.

Path to Production

Improve

- Can use model *as is*: **our approach does not rely on competition-specific assumptions**, and so it is **likely to generalize** out of sample and **perform well in realistic situations** with no further changes.
- [OPTIONAL] **Fine tune** the model with an augmented test set containing additional threats and body types, **further improving accuracy.**

Assess

- **Partner** with TSA, manufacturers to incorporate into software and **test.**
- **Deploy algorithm** in equipment and **conduct pilot trials** at select airports.
- **Collect feedback** from **operators**.

Deploy

- **Identify key learnings and modify / improve as necessary.**
- **Train personnel** and roll out nationwide.
- Develop processes to **continuously iterate and improve** software.