STEVENS
INSTITUTE OF TECHNOLOGY
1870

CNN Applications

FA690 Machine Learning in Finance

Dr. Zonghao Yang

2025 Spring

Case 1: Satellite Image and Poverty

Poverty Detection with Satellite Images

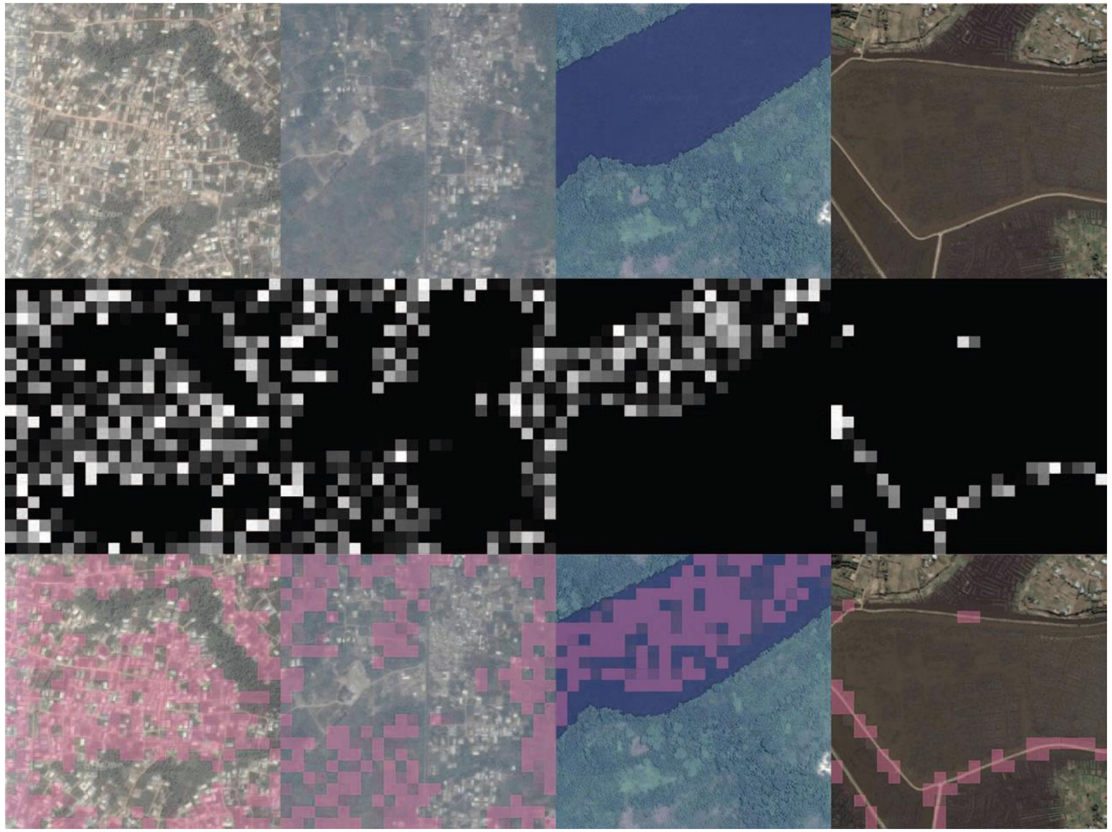
Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301), 790-794.

- Background
  - Reliable data on economic livelihood are scare in developing countries, which hampers efforts to study these outcomes and design policies that improve them
  - Traditional data collection methods (e.g., household surveys) are costly and difficult to scale
- Method: Train a convolutional neural network (CNN) to estimate economic livelihoods based on high-resolution satellite imagery
  - Nighttime light intensities: Nightlights serve as a noisy but globally consistent proxy for economic activity
  - Economic livelihoods: Consumption expenditures (World Bank's Living Standards Measurement Study) and asset wealth (asset index derived from the demographic and health surveys)
  - No human annotation
- The approach provides an accurate, inexpensive, and scalable method for estimating consumption expenditure and asset wealth using only publicly available data

Methodology

Transfer Learning: The authors use a multi-step transfer learning approach to train a CNN

- **Step 1**: Pretrain a CNN on ImageNet
- **Step 2**: Fine-tune the CNN to predict nighttime light intensities from daytime satellite imagery
- **Step 3**: Use the CNN to extract features from daytime imagery and train ridge regression models to estimate cluster-level economic outcomes (consumption expenditure and asset wealth)



a) Urban area b) Non-urban area c) Water body d) Road & transportation

Step 1: Pretraining

Pretraining a CNN on ImageNet, an image classification tasks for 1000 classes (animals, cars, etc.)

- Pretraining on ImageNet allows the CNN to learn low-level image features such as edges, corners, textures, and shapes
- These features are general and applicable to many visual tasks, not just the specific task of predicting economic outcomes
- **Why Pretrain?** The target dataset (in this case, satellite imagery) has limited labeled data

Step 2: Fine-tuning

The pretrained CNN is fine-tuned to predict nighttime light intensities from daytime satellite imagery

- Nighttime light intensity (measured by satellites) is a widely used proxy for economic activity
  - Areas with higher economic activity tend to have brighter nightlights due to electricity usage, urban development, and infrastructure
- Process:
  - The CNN is trained to take daytime satellite images as input and predict the corresponding nighttime light intensity for that location
  - During this step, the CNN learns to summarize the high-dimensional daytime imagery into a lower-dimensional set of features that are predictive of nighttime light intensity
  - These features include semantically meaningful patterns such as urban areas, roads, water bodies, and agricultural land, even though the model is not explicitly told to look for these features
- The fine-tuned CNN acts as a feature extractor, producing a concise representation (feature vector) of each daytime image that captures information relevant to economic activity

Step 3: Predicting Economic Outcomes

Use the features extracted by the CNN to predict cluster-level economic outcomes, such as average household consumption expenditure and average household asset wealth

- Feature Extraction: The fine-tuned CNN provide lower-dimensional representations of the satellite images, capturing information about the landscape that is correlated with economic activity.
- Ridge Regression: The extracted features are used as inputs to a ridge regression model, which is trained to predict economic outcomes.
- Prediction targets: Cluster-level consumption expenditure and asset wealth
  - The model predicts economic outcomes at the cluster level, which is the smallest geographic unit for which latitude and longitude data are available in the public-domain surveys
  - The predictions are validated against survey data using cross-validation to ensure the model's accuracy and generalizability

The Importance of Transfer Learning

- The main **challenge** in this setting is the scarcity of labeled data for economic outcomes. Detailed household surveys are expensive and time-consuming to conduct, and they typically cover only a small fraction of the population.
- The authors use **transfer learning** to overcome this challenge. By first training the CNN on a large, unrelated dataset (ImageNet) and then fine-tuning it on a proxy task (predicting nightlights), the model can learn useful features even with limited labeled data on the target task (predicting economic outcomes).
- Transfer learning allows the model to leverage knowledge from one domain (general image classification) and apply it to a different but related domain (economic prediction), significantly improving performance in data-scarce settings.

Case 2: Evaluate Delinquency Risk through Real-Time Video Analysis

Evaluate Delinquency Risk

Chang, X., Dai, L., Feng, L., Han, J., Shi, J., & Zhang, B. (Forthcoming). A Good Sketch is Better than A Long Speech: Evaluate Delinquency Risk through Real-Time Video Analysis. *Review of Finance*.

- Research question: Do borrowers' microfacial expressions during the loan application process predict their likelihood of loan delinquency?
- **Happiness** Borrowers who exhibit more micro-facial expressions of happiness are less likely to default on their loans.
- **Fear** A higher frequency of fear-related expressions correlates with increased delinquency risk.
- The paper demonstrates that incorporating facial expression data alongside traditional credit scores improves the accuracy of delinquency prediction models.

An Illustrative Frame in Loan Application Video

This figure presents one frame example of a loan applicant's video. They identify and differentiate the applicant from the staff and other people in the shopping mall, and then apply the machine learning video analysis to capture the applicant's facial expressions during the loan application process.

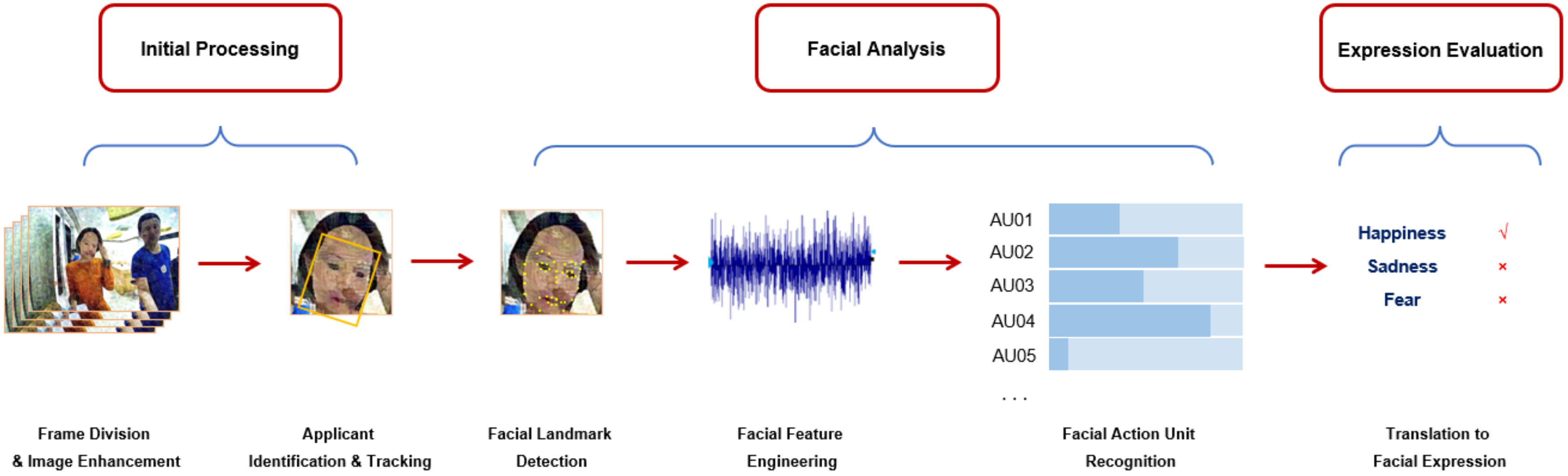
Methodology

Openface: an open source **facial behavior analysis toolkit**
T Baltrušaitis, P Robinson… - 2016 IEEE winter …, 2016 - ieeexplore.ieee.org
… core technologies used by **OpenFace** for **facial behavior analysis** (see Figure 2 for a **summary**). First, we provide an explanation of how we detect and track **facial** landmarks, together …
☆ Save 99 Cite Cited by 3930 Related articles All 17 versions

OpenFace 2.0 is an open-source facial behavior analysis toolkit developed by researchers at Carnegie Mellon University and Microsoft.

- Video preprocessing
  - Frame extraction: Each video is split into individual frames at a rate of 10-30 frames per second
  - Applicant identification: Since videos may contain multiple people (e.g., loan applicants and bank staff), OpenFace identifies the loan applicant as the person appearing in the center of the first video frame
- Facial landmark detection
  - Uses a **Convolutional Experts Constrained Local Model (CE-CLM)** for facial landmark detection
  - Identifies **68 key facial points** (e.g., corners of the mouth, eyebrows, nose tip, eyes) in each frame
  - Tracks facial changes over time to capture subtle expressions.
- Examples
  - Raised **cheeks and lip corners** → Happiness
  - Raised **eyebrows, widened eyes** → Fear

Machine-Learning Based Video Analysis



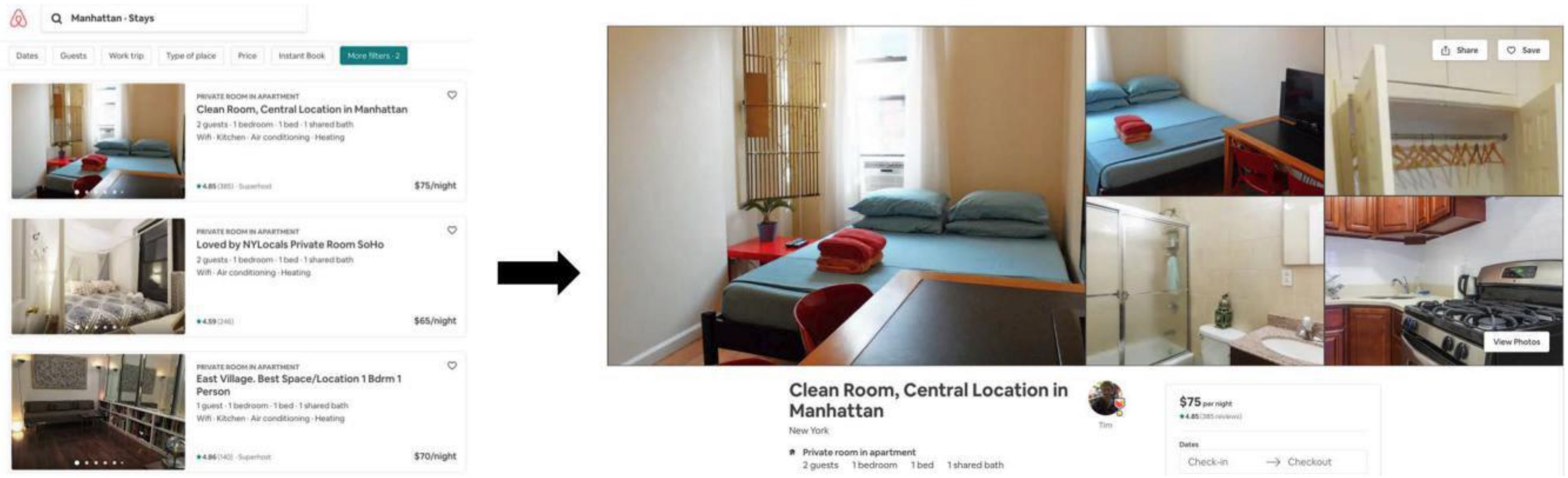
Facial Expression

Happiness

Sadness

Fear

Case 3: Airbnb Image Quality



Airbnb Image Quality

Li, H., Simchi-Levi, D., Wu, M. X., & Zhu, W. (2023). Estimating and exploiting the impact of photo layout: A structural approach. *Management Science*, 69(9), 5209-5233.

- Airbnb hosts use photos to showcase their properties, but the quality, content, and order of these photos can significantly influence customer decisions
- The paper quantifies the impact of photo layout (i.e., the type of room featured, photo quality, and display order) on customer booking decisions and to optimize the photo layout to maximize rental demand and revenue
  - The authors use *ResNet-50* to evaluate photo quality and classify room types (bedroom, living room, outside, toilet, kitchen)
- Main results
  - **Cover Photos**: The cover photo has a significantly larger impact on customer decisions than non-cover photos. A high-quality bedroom cover photo leads to the largest increase in demand.
  - **Room Types**: Bedroom photos are the most effective for cover images, while living room photos, commonly used by hosts, are less impactful.

CNN for Image Analysis

The authors use ResNet-50, a convolutional neural network (CNN), to analyze the photos posted by Airbnb hosts.

- Two Separate Models:
  - Image Quality Model: A regression model that assigns a quality score (1 to 7) to each photo based on its visual attractiveness. The model is trained on a subset of 4,000 images, each rated by four human subjects.
  - Room Type Classification Model: A classification model that categorizes photos into five room types: bedroom, living room, outside, toilet, and kitchen. The model achieves an accuracy of 84%.
- Transfer Learning: ResNet-50 is pretrained on ImageNet, a large image classification dataset, and fine-tuned for the specific tasks of quality scoring and room type classification.
- Data Augmentation: To prevent overfitting, the authors apply data augmentation techniques such as random rotation, horizontal flips, and random crops.
- Integration with economic modeling: The extracted photo features (quality scores and room types) are used as inputs in a Pairwise Comparison Model (PCM) to estimate the impact of photo layout on customer booking decisions.

STEVENS
INSTITUTE OF TECHNOLOGY
1870

THANK **YOU**

Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030