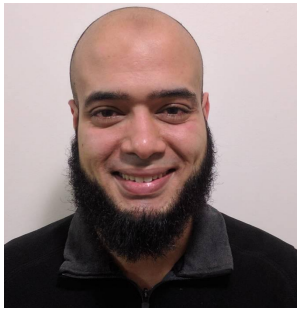
Machine Learning

Unlimited Data Variance Challenge

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)



Data Variance

- In Computer vision, we learned so far that we can collect fine-grained data collection
  - For example, identify all possible scenarios to put your hands on wheel
  - This is a great but long process
- We also learned to think about the variables changing
  - Gender: men vs women
  - Skin Color
  - Accessories and clothing
- This type of variance can be handled to a large extent with process we learned
- However, sometimes this is not enough to provide wide coverage

- With gloves, there are unlimited or huge possible colors, textures and styles!

- With steering wheels, there are a huge possible colors combinations!

Tackling the challenge

- When you have a variance in a feature, you have 2 paths:
  - Make the model invariant to it: prepare data with all possible diversity
    - Cool, but you put effort in collecting or **augmenting/synthesizing** the data
  - Make the data normalized: find a way to map all features to limited range/diversity
    - Great, but not always possible
- For the gloves/wheel, discuss the 2 solutions

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."