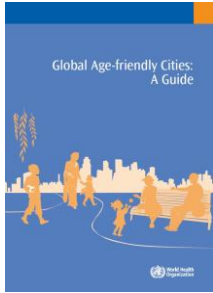幻灯片 1

Hello everyone; I am very honoured to have this **opportunity to share** with you the project that I am working on. Our research **topic** is about "Using artificial intelligence wisely for deep mapping the urban perceptions of Age-friendliness."

The World Health Organization has identified the need to **consider older adults in urban design** as a **key finding** of its **work on Age-Friendly Cities**. This was based on data from focus groups conducted in **33 cities around the world**. However, **despite** this important finding, little research has been published **linking age-friendly communities as a whole with the well-being of older residents**.
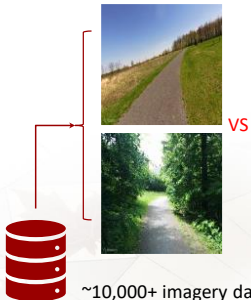
Several years ago, Dr. Kristjansson and her team at the University of Ottawa conducted **relevant research on age-friendly communities**. They found that **greenspace and whitespace** were two key measures influencing age-friendliness and well-being. Their study was carried out in **eight cities across Canada**. Building upon their work, we have focused our research on **the assessment of greenspaces in Ottawa**, using **street view images** photographed in **the summertime** for greenspace quality assessment.

We have done previous work **on collecting the data by collaborating with** a volunteer group called 'Snow Mole' from Ottawa. We together **established four perceptual dimensions** of greenspace quality based on **four questions**: "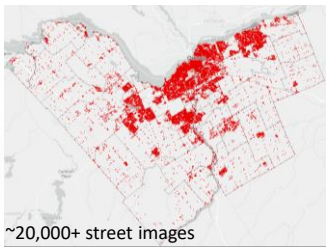In which environment do you think has a strong connection to nature?", "Which environment do you find more beautiful or appealing?", "In which environment would you feel safest?" and "In which environment do you feel you would best be able to do the activities?". **Over 10,000** street images have been rated for each question **in the form of random pair duel**.

Our aim is to create **detailed maps** across Ottawa, with each **street and park ranked** along each age-friendly dimension. However, given the large number of images to be predicted, it is **impracticable** to ask people to assess and rank them **manually**, as this would require us to train hundreds of people capable of completing such a task.

Therefore, we have developed an approach using a convolutional neural network, a **popular method** in computer vision **for mapping perceptual constructs** associated with street images.

During our data collection process, participants sat in front of a computer and were presented with a web page displaying two images **side-by-side**. They then chose **either the left or right** image based on a qualitative question, such as which one exhibits a stronger connection to nature to them. We only recorded the **identifier** of the chosen images and the **winner** of the comparison **in tables**.
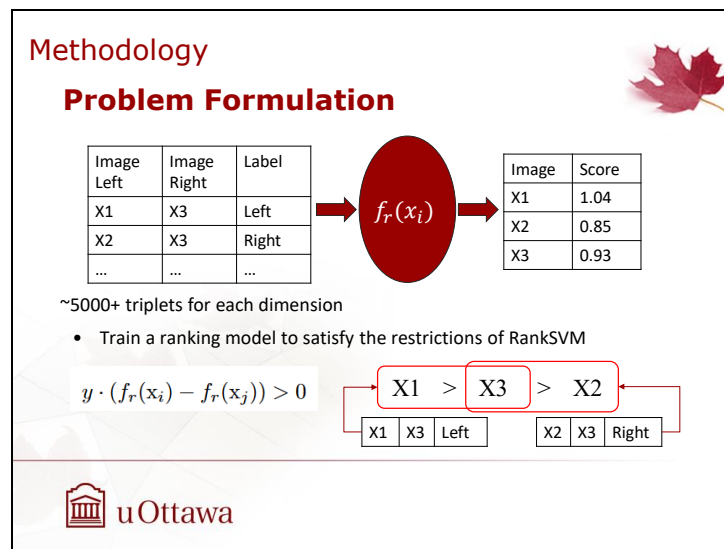
Our dataset comprises of over 5,000 records of **pairwise comparison** for each perceptual dimension. Our aim is **not just to** identify the winner of each duel, but to create maps that exhibit the **qualitative characteristics** of age-friendliness. Hence, we strive to find a ranking function or model that can **quantify the duel results**. To achieve this, we have formulated our questions **in accordance with** the RankSVM constraint, which is a **well-established algorithm** that transforms a ranking learning problem into an optimization problem with a set of constraints.
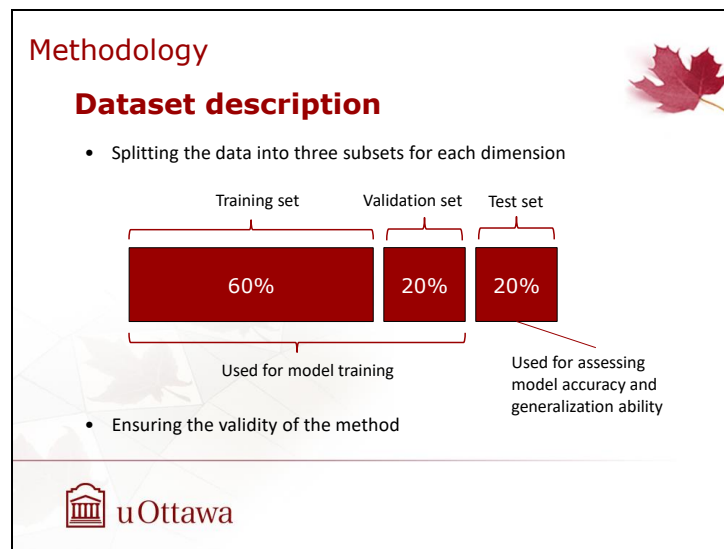
To ensure the **validity** of our method, we have **partitioned** our dataset into three subsets for each dimension: a training set for **training the neural network**, a validation set for evaluating the **prediction performance** and **adjusting model hyperparameters** as needed, and a test set that is **never seen** during the model training process and is only **used once** to **independently assess the model's accuracy**.
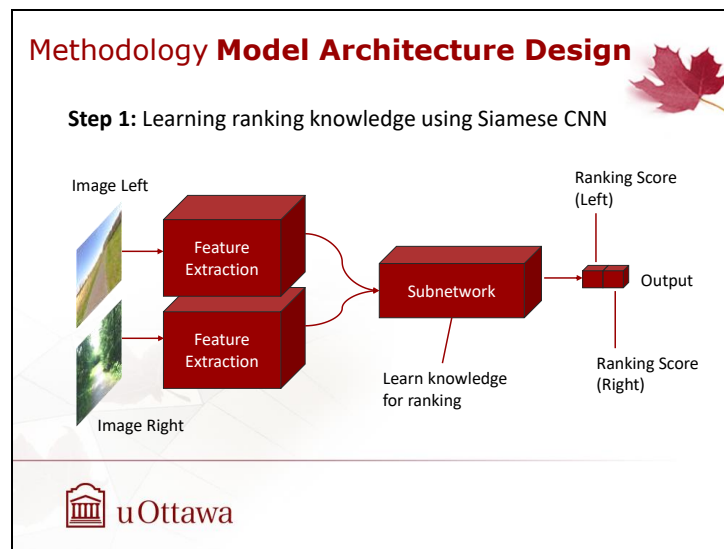
Our approach involves designing a model that learns to rank the training data **using a Siamese CNN architecture**. The model consists of two **identical, disjoint branches** that **extract visual semantic features** from the two images in a pair. These feature extractor layers are then concatenated and followed by **a subnetwork that learns the ranking knowledge**. Finally, the model **outputs the ranking score** of the two input images.
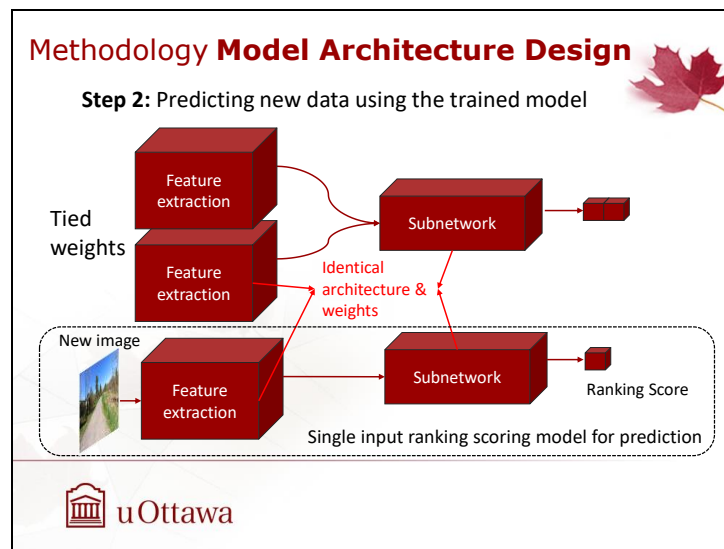
Once the model is well-trained, we build another **ranking scoring model** for **a single input** using **the same** feature extraction and the subnetwork architecture with **identical trained weights**. We then use this ranking scoring model **to predict new data**.
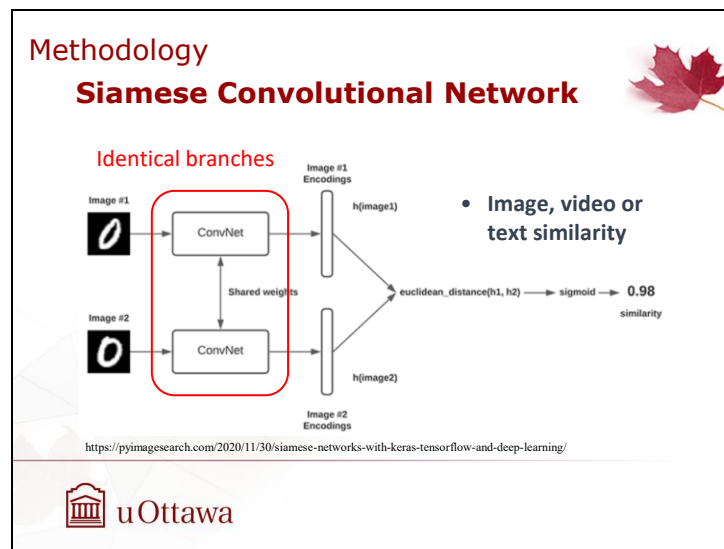
A Siamese Neural Network is **a powerful tool** for **learning joint feature representations** in a wide range of applications, such as **image, video or text similarity.** This type of network has two **identical branches with the same configuration, parameters and weights**, and the parameter updating is mirrored across both subnetworks. Therefore, we think it is **ideally adapted** to our problem formulation
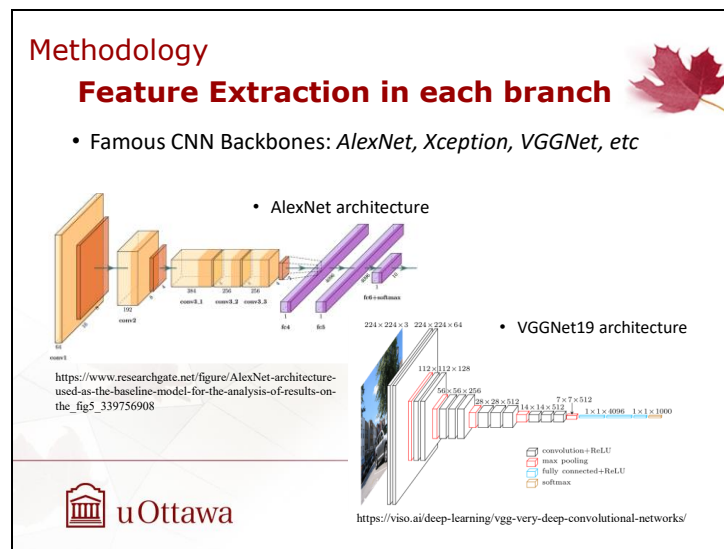
The process of feature extraction is **critical in image analysis**, as it allows for **the conversion of raw image data into multi-level representations** that can be used for a variety of tasks. **Convolutional neural networks** have proven particularly effective for this purpose, as they can **automatically learn complex features** in images. In our research, we experimented with several open-source CNN backbone that have demonstrated **remarkable performance on classification tasks on the ImageNet dataset**, which is a dataset with over one million images. The backbones we chose include AlexNet, Xception, VGGNet and others.
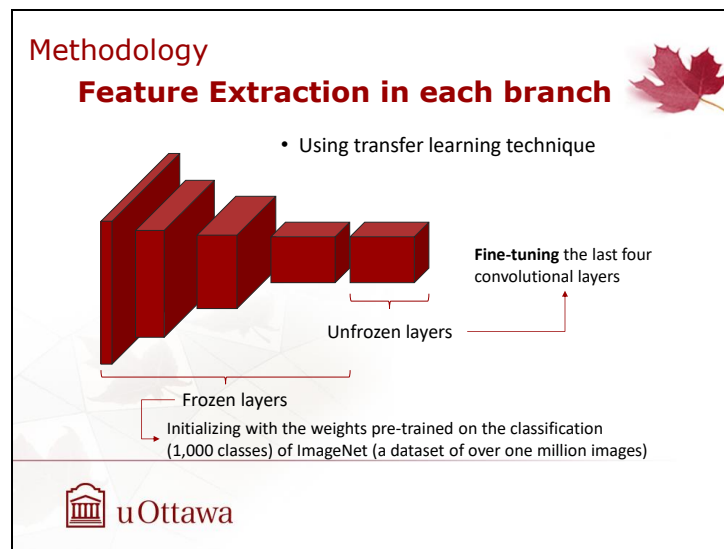
However, due to the small dataset we collected, there is **a risk of overfitting and weak-generalizing ability**. To mitigate these risks, we employed a **transfer learning technique** for the backbone. Specifically, we **initialized the weights** with pre-trained parameters on the ImageNet dataset. Additionally, we made **slight adjustments** to the last several layers of the network **to ensure its relevance** to our problems and datasets at hand.
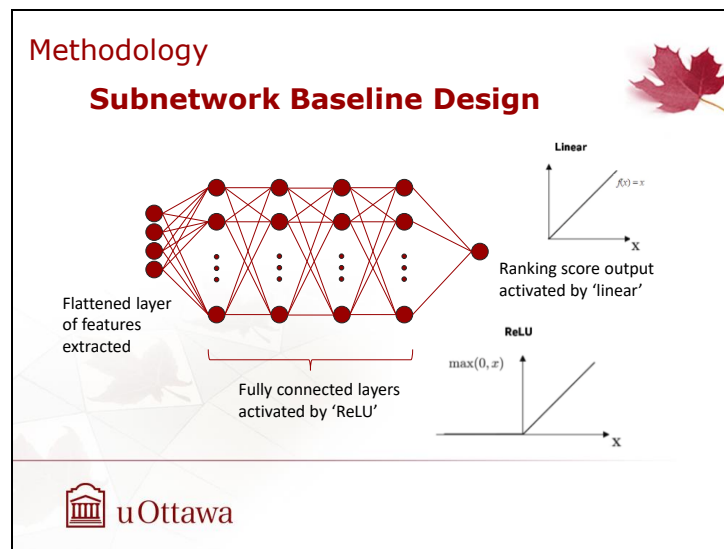
To model the ranking function, we have designed a subnetwork that consists of **a sequential layer group of fully connected layers**. In detail, the neurons in each layer are activated **by the ReLU function** and the last output layer is activated by **a simple linear function to yield ranking scores**.
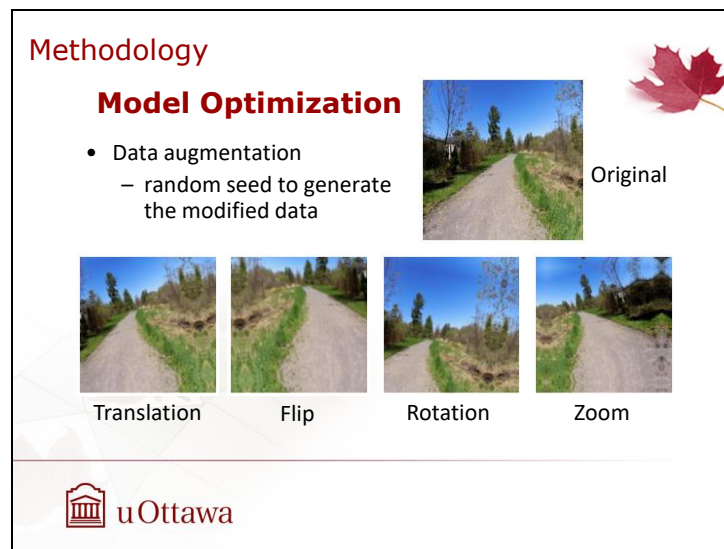
To better leverage limited data, we employed **four types of data augmentation techniques**, including **translation, flipping, rotation, and zoom**. Translation involved **shifting the image in different direction**s, while flipping involved **horizontally flipping** the image. Rotation **entailed rotating the image by a certain degree**, and zooming involved **scaling the image**. To ensure that our augmentation was performed **consistently**, we utilized **a random seed** to generate the modified data.
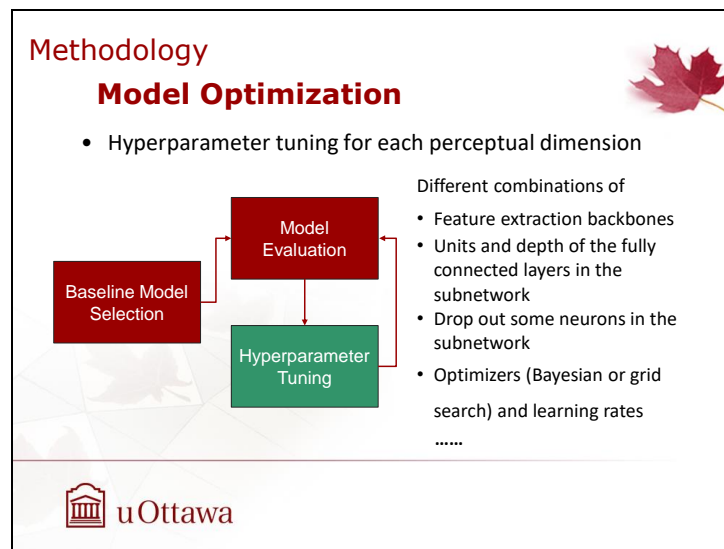
Hyperparameter tuning is **another important means** to optimize the model performance. In our study, we have thoroughly **tuned various hyperparameters** to ensure that our model **performs at its best**. We have experimented with different **CNN backbones, fully connected layer configurations, activation functions, dropout rates, optimizers, learning rates**, and more. Through a systematic process of **trial and error**, we were able to **find the optimal combination of hyperparameters** that produced the best results **on the validation sets**.
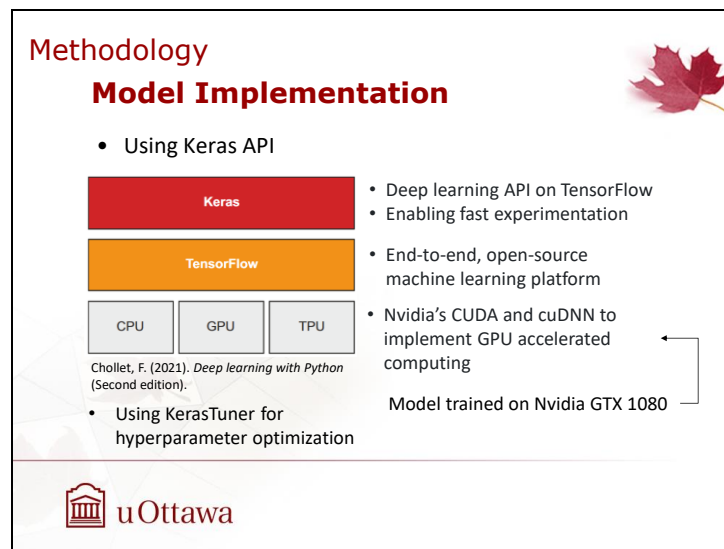
Our model is implemented using the Keras API, an **open-source** deep learning API built **on top of TensorFlow** that enables **efficient experimentation**. We leverage the computational power of an Nvidia GTX 1080 GPU, which is capable of **accelerating computations with the support of CUDA and cuDNN frameworks**, for efficient model training.
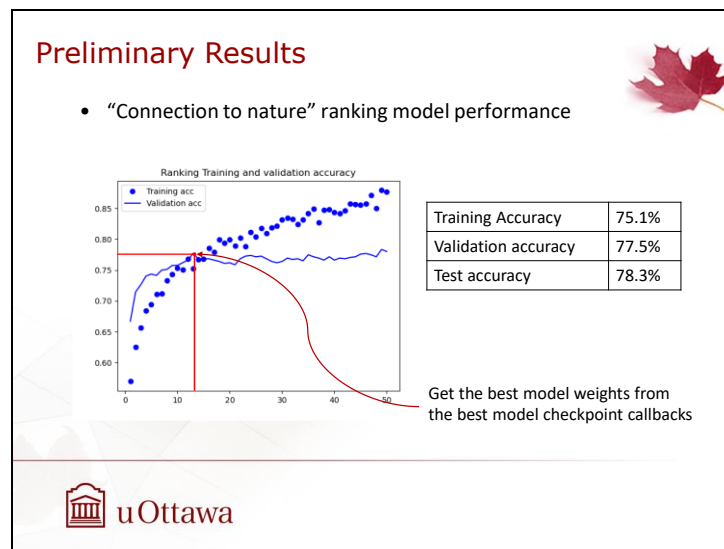
We are pleased to report that we have obtained **promising preliminary results** from our training on the 'Connection to nature' dimension. Although **overfitting is a concern** since the gap between training and validation accuracy is increasing, the model performance is **generally satisfactory** if we stop at the epoch when the highest validation accuracy occurs. As shown in the slide, our ranking model has achieved a **training accuracy of 75.1%** and a **validation accuracy of 77.5%.** Moreover, the model's test accuracy is **even higher** at 78.3%, indicating that the model has **a good generalization ability**.
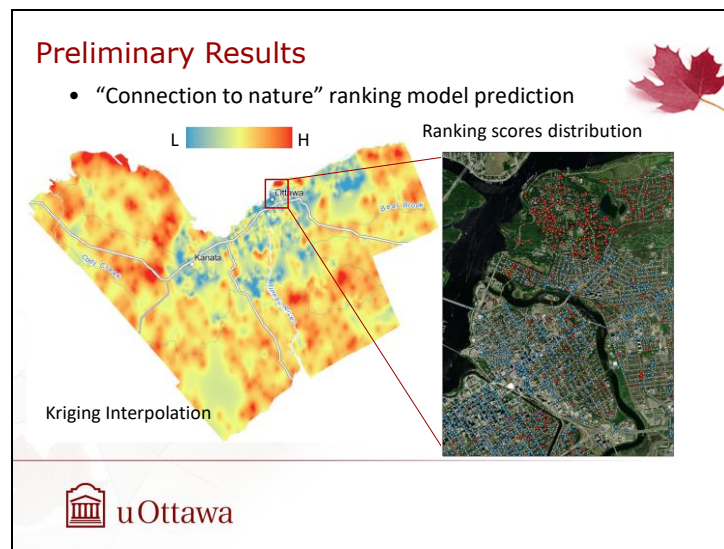
Subsequently, we applied our trained model to **predict the ranking scores of new images** in Ottawa, and we constructed **an interpolation map for evaluation purposes**. The map depicts areas that convey a strong sense of 'connection to nature' **in blue**, while areas that do not **in red**. Notably, the blue regions tend to **correspond to areas** with **lower** vegetation cover and **higher** human-made structures. Upon **closer examination** of a specific area in the downtown region on the right, we found that the high predicted ranking score areas that are **colored red** and the low predicted ranking score areas that are **colored blue** are **consistent** with our **visual analysis** of the results from the satellite background.
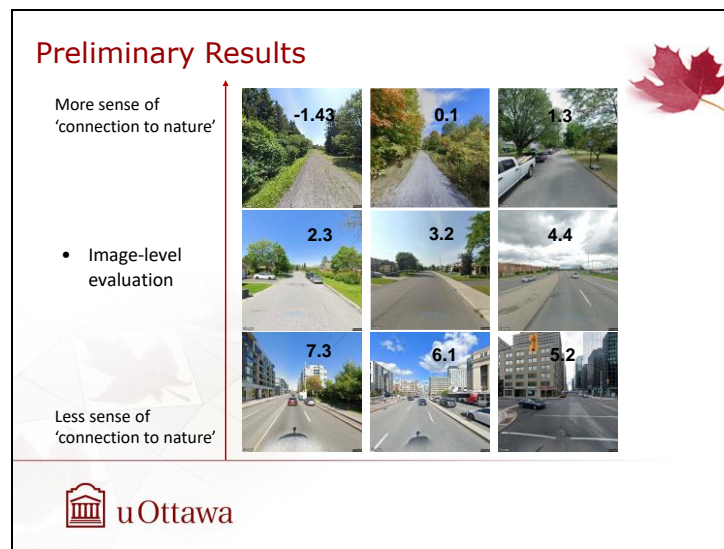
Furthermore, we conducted an **image-level evaluation** of our model's performance. We performed a **visual comparison** of the model's predicted ranking scores with the selected images, as shown in the slide. The **comparison confirmed** that the model's performance was **indeed satisfactory**, as the predicted scores **aligned well with human perception**.
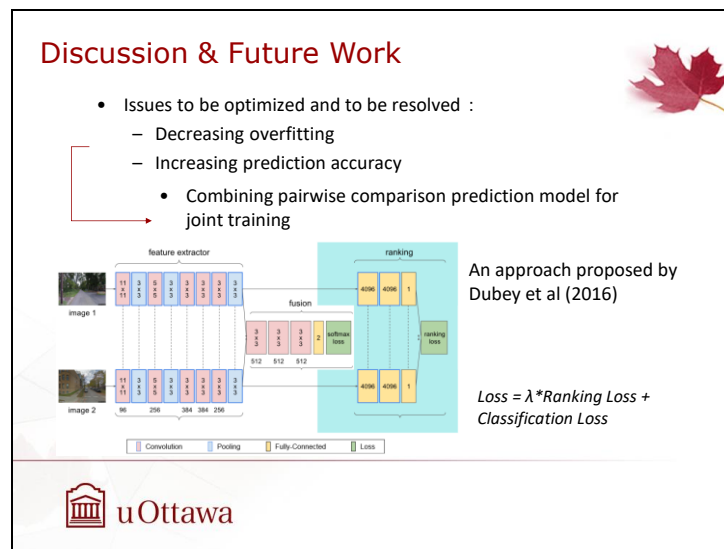
Our study is still in its early stages, and there are **several issues** we need to address to **improve** the model's accuracy and **reduce** overfitting. In reference to existing literature, one approach that has shown promise is **the combination** of **building a** pairwise comparison **model** and **a** ranking **model** simultaneously **for joint training**, as proposed by Dubey et al in 2016. We plan to **adopt** this approach in our research to see if it **improves** our prediction accuracy.
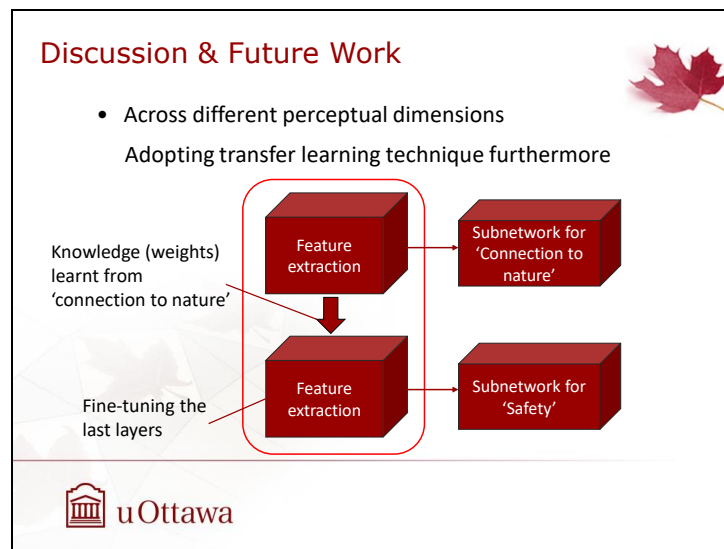
Moving forward, we intend to train models for the three **remaining** perceptual dimensions. One technique we will consider is transfer learning **across different dimensions**. For instance, we may transfer the **weights learned from** the 'Connection to Nature' ranking model **to** the 'Safety' ranking model, which we presume could potentially **yield higher accuracy than** just using the weights pre-trained on ImageNet **alone**.

Thank you for your listening.