Adaptive Graph Attention Network in Person Re-Identification

Presenter: Le Dinh Duy
Roll ID: HE130655
Computer Science Student, FPT University, Hanoi, Vietnam

ABOUT ME

Computer Science Student

FPT University, Hanoi

Email: duyldhe130655@fpt.edu.vn

INTRODUCTION

+ Person re-identification has a huge potential in applications related to video surveillance.
+ There are several issues the re-id individual has to face.

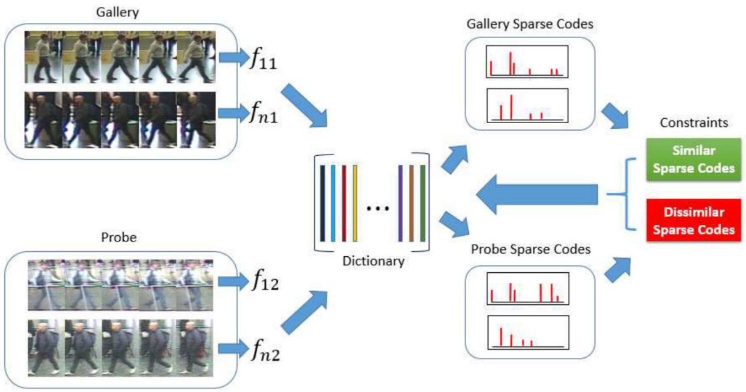
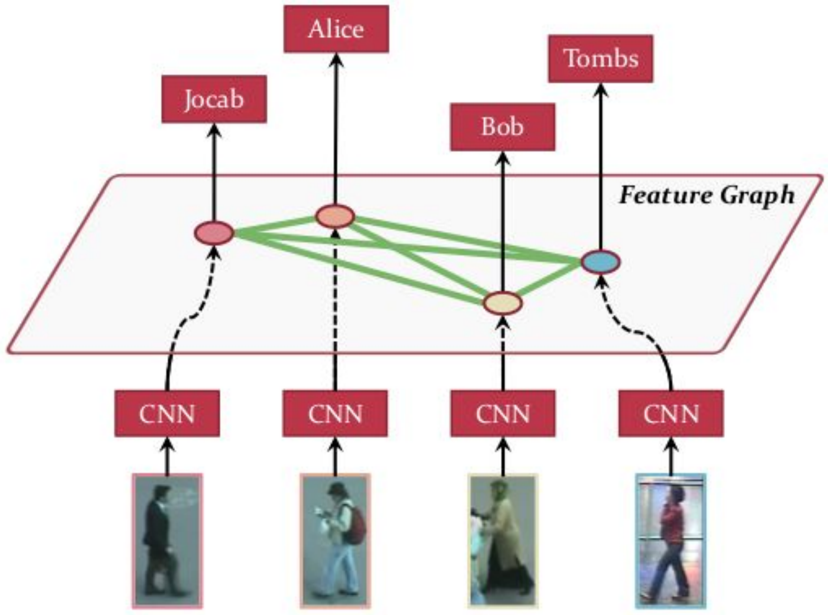


Figure 2: A visual summary of our training process. Given image sequences in both gallery and probe camera views for n persons, we first compute their representative feature vectors. We then iteratively train a discriminative viewpoint invariant dictionary by imposing explicit constraints on the corresponding gallery and probe sparse codes.

INTRODUCTION

The target of this research is to improve the learning efficiency of the graph approach with the re-id task of the application of the contextual.



AGENDA

- INTRODUCTION
- RELATED WORD
- METHODOLOGY
- RESULTS ANALYSIS
- CONCLUSIONS

RELATED WORD

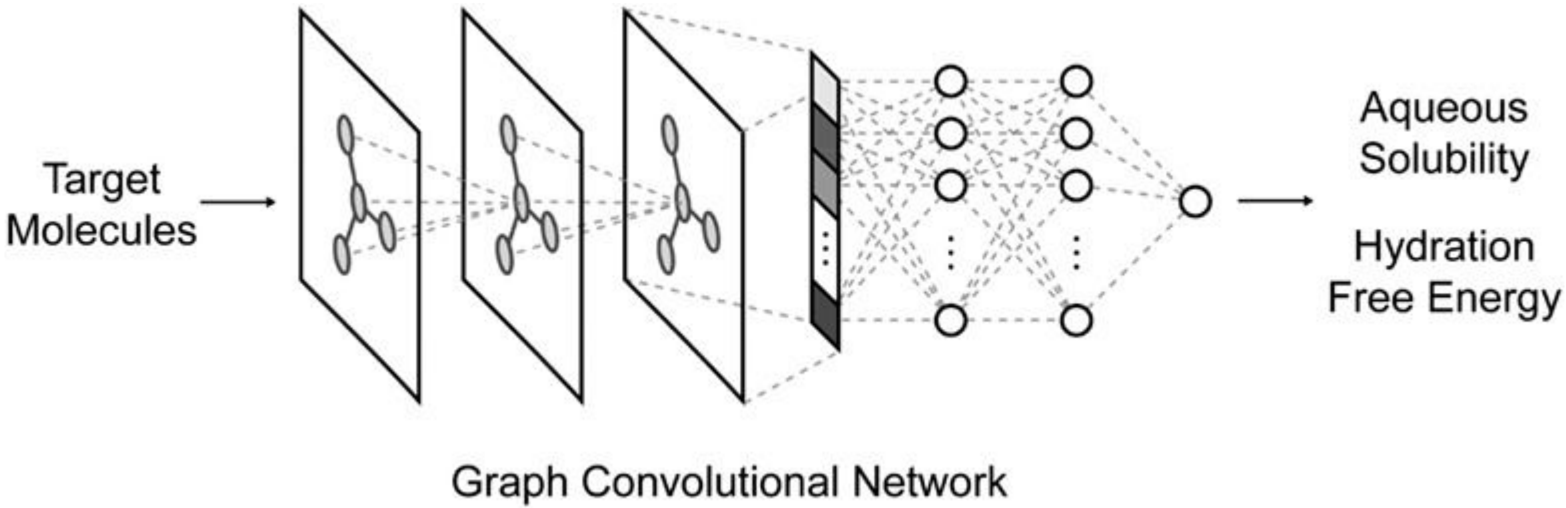


Illustration of the graph convolutional network for prediction of aqueous solubility from molecular structure.

RELATED WORD

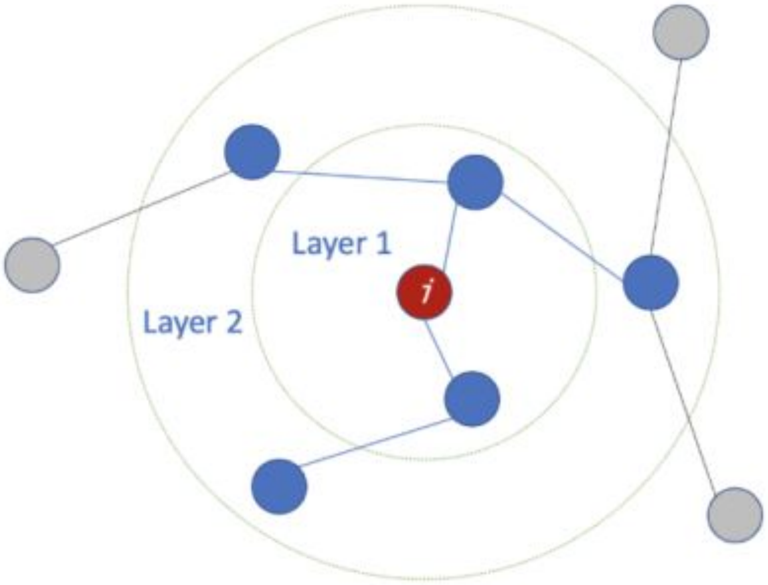


Figure 2.1: Example: Gathering info process with 2 layers of target node i.

RELATED WORD

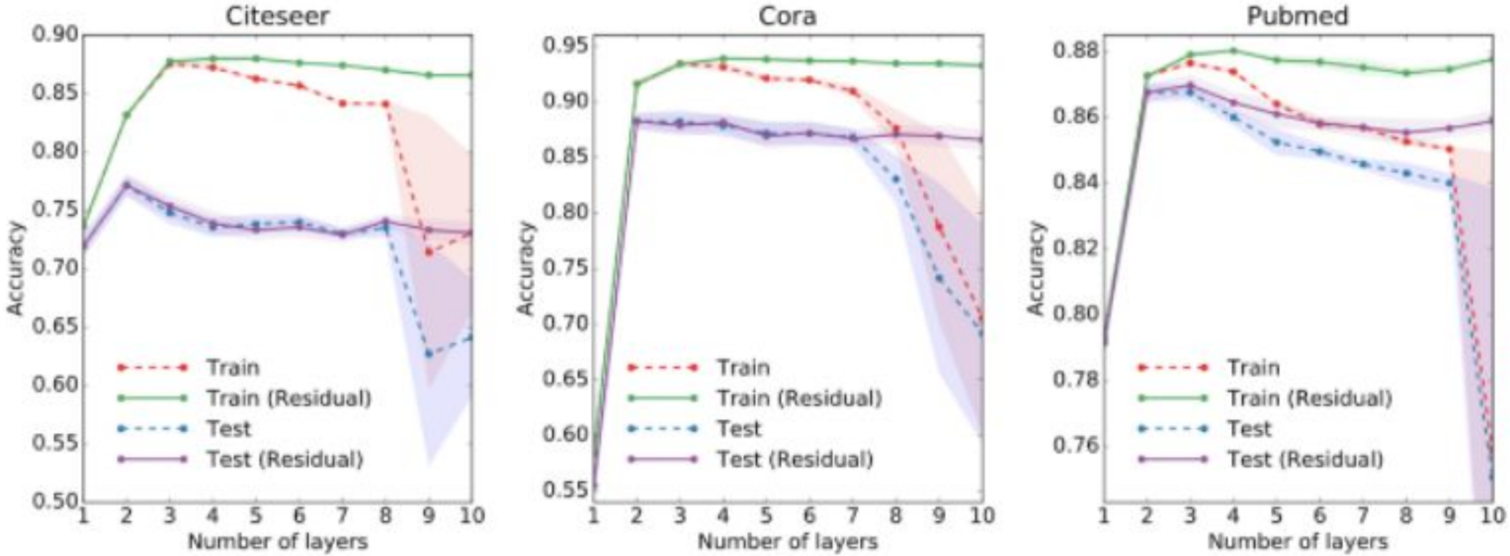


Figure 2.2: Performance over layers. Picture from the paper [2]

RELATED WORD

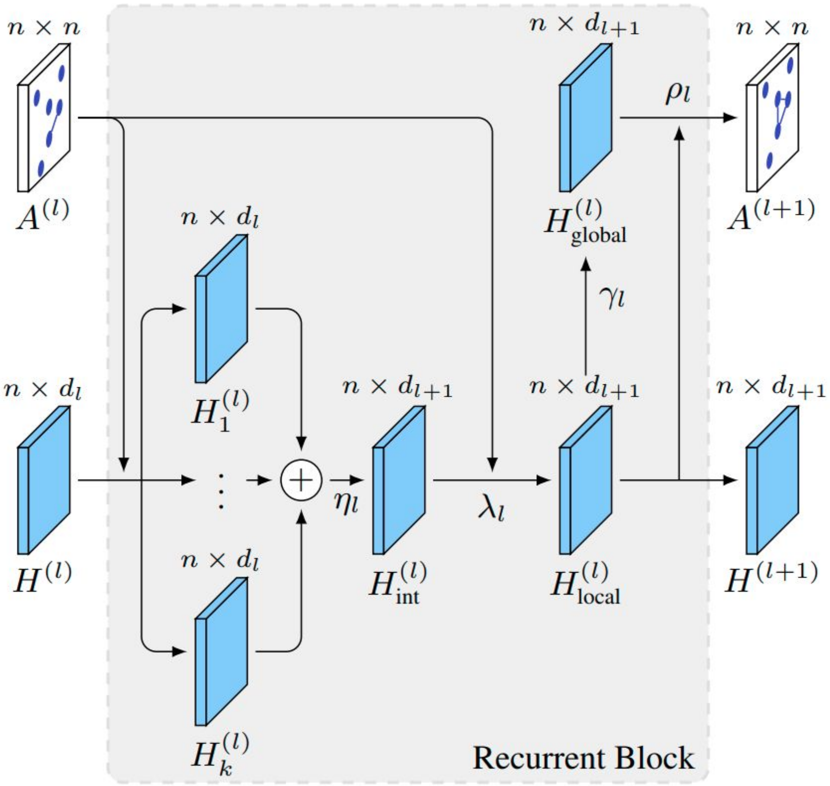


Figure 2.4: The author proposed method is a recurrent block. They create a set of node embeddings $\{H_i^{(l)}\}_{i=1}^{k}$ that are later combined to produce an intermediary representation $H_{int}^{(l)}$. Then use the updated node information with the adjacency information to produce a local embedding of the nodes information $H_{local}^{(l)}$ thats is also the output $H^{(l+1)}$. Also broadcast the information of the local embedding to produce a global embedding $H_{global}^{(l)}$. And combine the local and global embeddings to predict the next layer adjacency $A^{(l+1)}$[3].

RELATED WORD

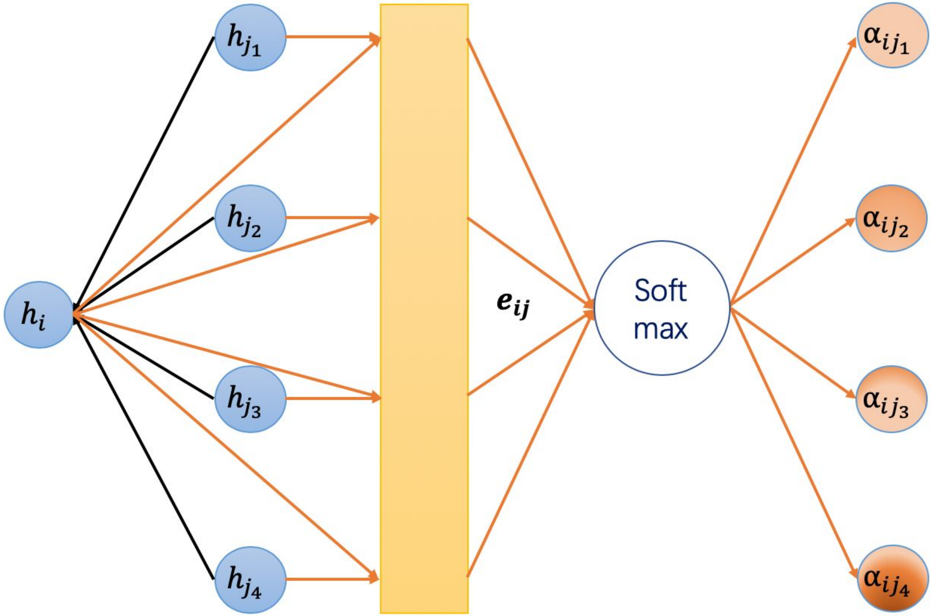
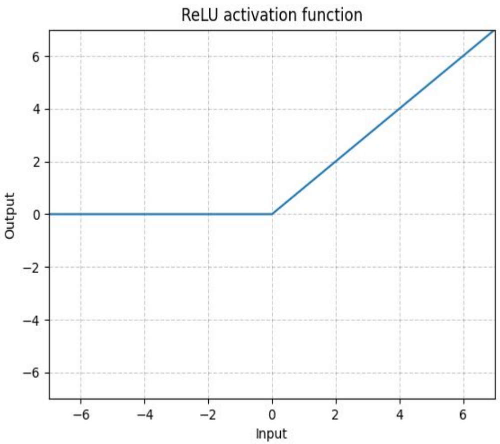
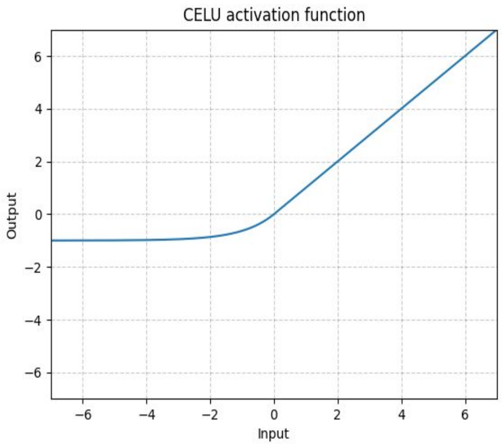


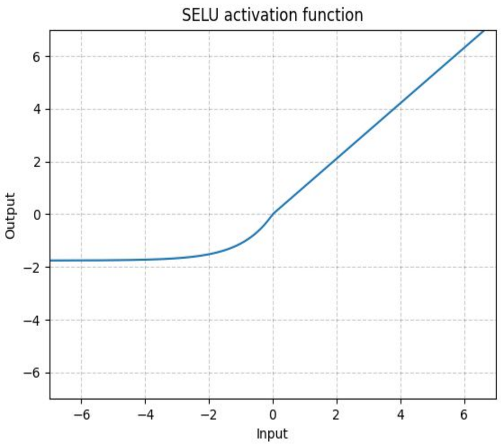
Figure 2.3: The attention mechanism $\vec{a}(W_{\vec{h}i}, W_{\vec{h}j})$ employed by GAT model, parametrized by a weight vector $\vec{a} \in \mathbb{R}^{2F'}$, applying a LeakyReLU activation.

RELATED WORD

The activation functions







$$RELU(x) = (x)^{+} = \max(0, x)$$

$$SELU(x) = \begin{cases} \lambda x & \text{if } x \geq 0 \\ \lambda \alpha \left(\exp\left(x\right) - 1\right) & \text{if } x < 0 \end{cases}$$

$$CELU(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha \left(\exp\left(\frac{x}{\alpha}\right) - 1\right) & \text{otherwise} \end{cases}$$

RELATED WORD

Cosine embedding loss

$$\mathcal{L}(x,y) = \begin{cases} 1 - \cos(x_1, x_2) & \text{if } y = 1, \\ \max(0, \cos(x_1, x_2)) & \text{if } y = -1, \end{cases} \tag{2.12}$$

Graph learning loss

$$\mathcal{L}_{GL} = \frac{1}{N^2} \sum_{i,j=1}^{N} \exp\left(A_{i,j} + \eta ||v_i - v_j||_2^2 \right) + \gamma ||A||_F^2 \tag{2.13}$$

RELATED WORD

Regularization

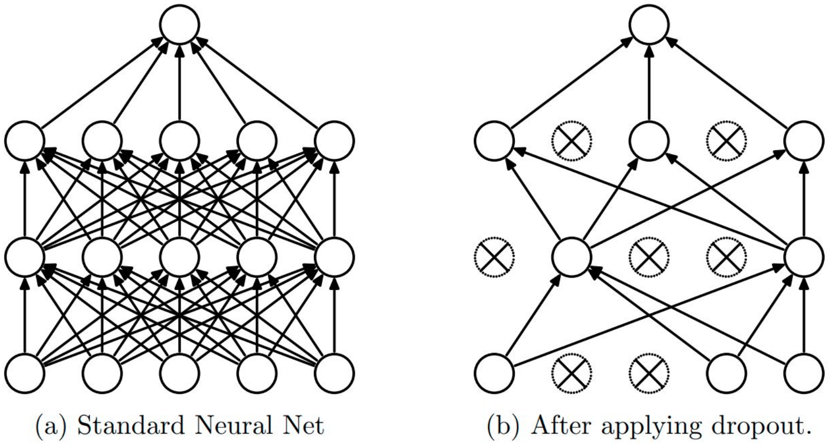


Figure 2.9: Dropout Neural Net Model. Left: A standard neural net with 2 hidden layers. **Right**: An example of a thinned net produced by applying dropout to the network on the **left**. Crossed units have been dropped..[4]

RELATED WORD

Regularization

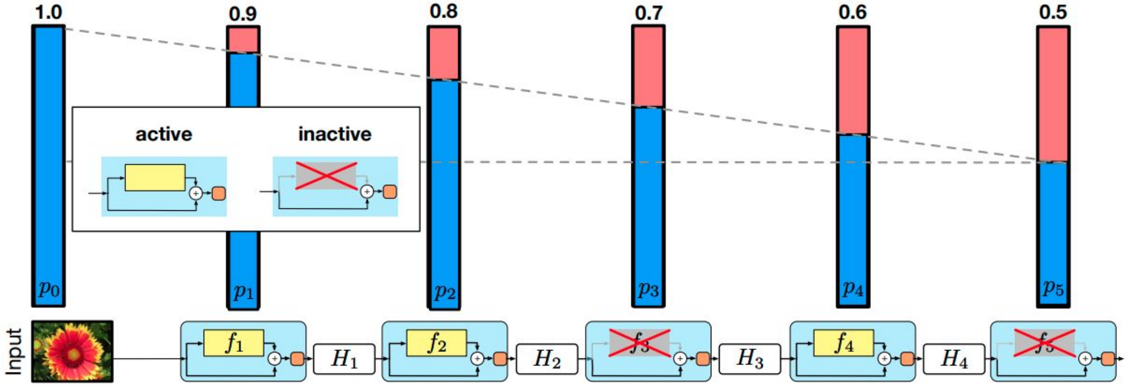


Figure 2.10: The linear decay of p_l illustrated on a ResNet with stochastic depth for $p_0 = 1$ and $p_L = 0.5$. Conceptually, we treat the input to the first ResBlock as H_0, which is always active.[4]

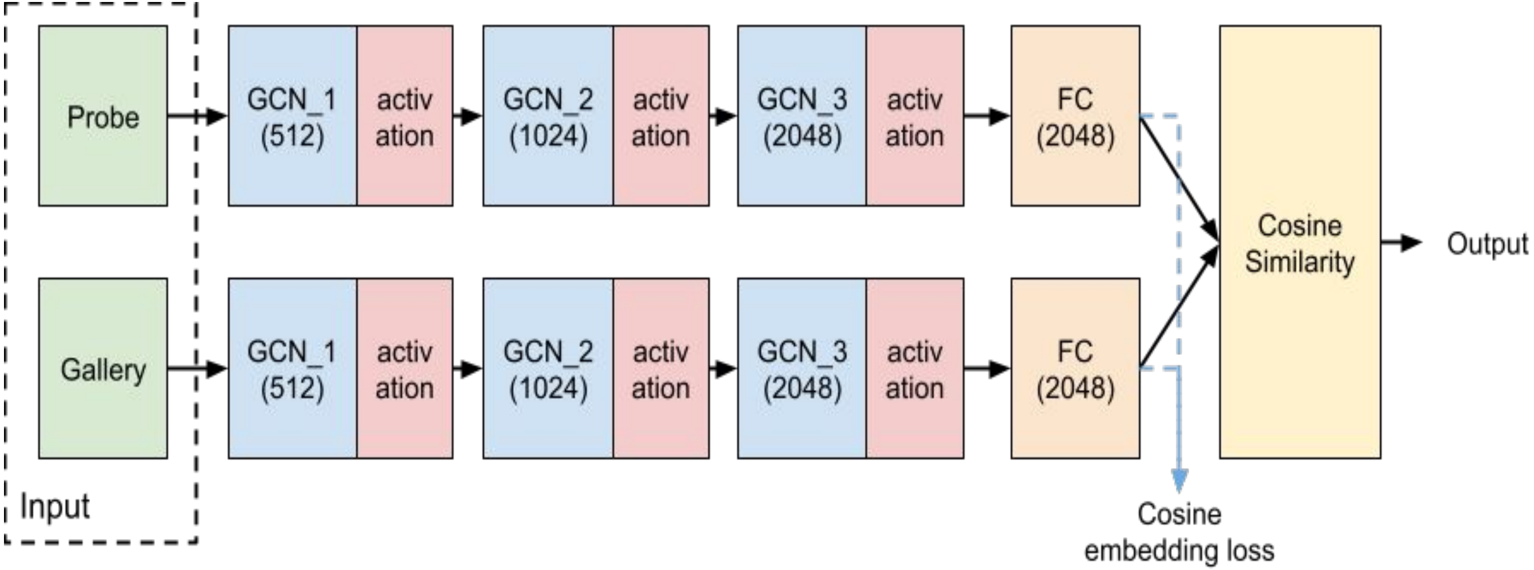
METHODOLOGY

The attention central node matrix

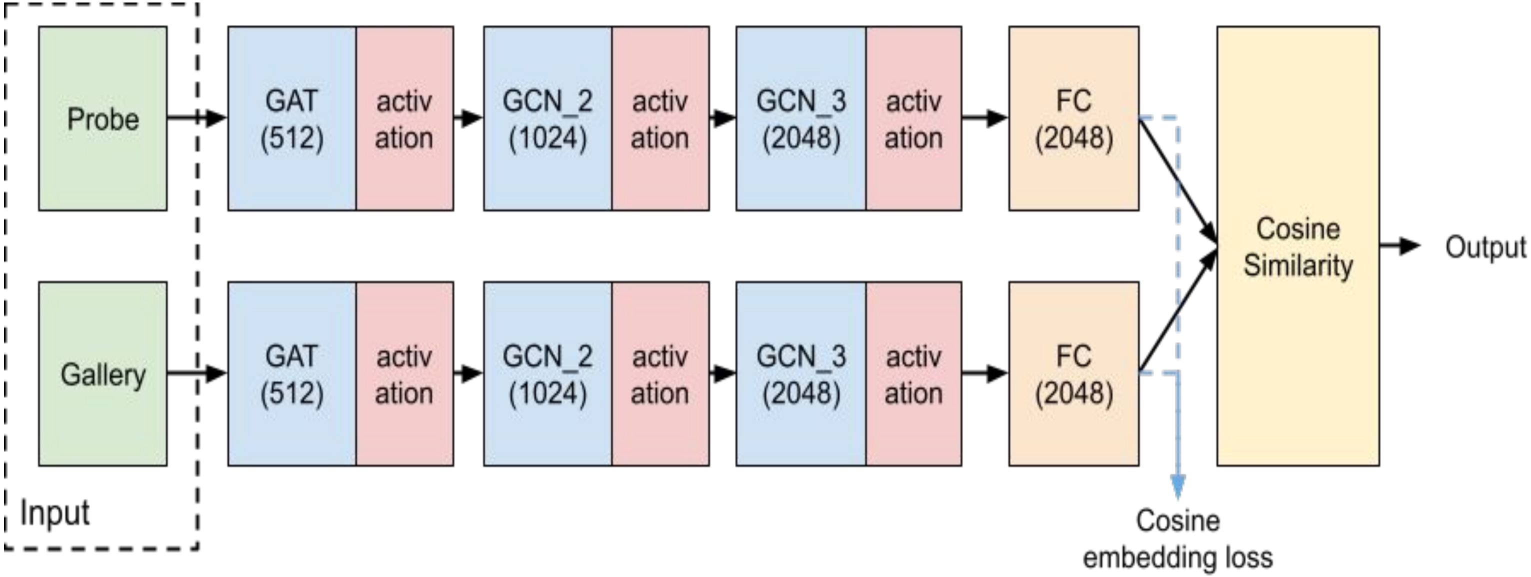
$$\tilde{A}_{i,j} = \begin{cases} \sqrt{2/N} & if \quad j = 1 \\ \sqrt{N/2} & if \quad i = 1 \\ 1 & if \quad i = j \\ 0 & otherwise \end{cases}, \tag{3.1}$$

We use A ∈ R(N×N) to denote the adjacent matrix associated with graph G. We assign the target node as the first node in the graph, and normalize by assume above, then add self-loop,

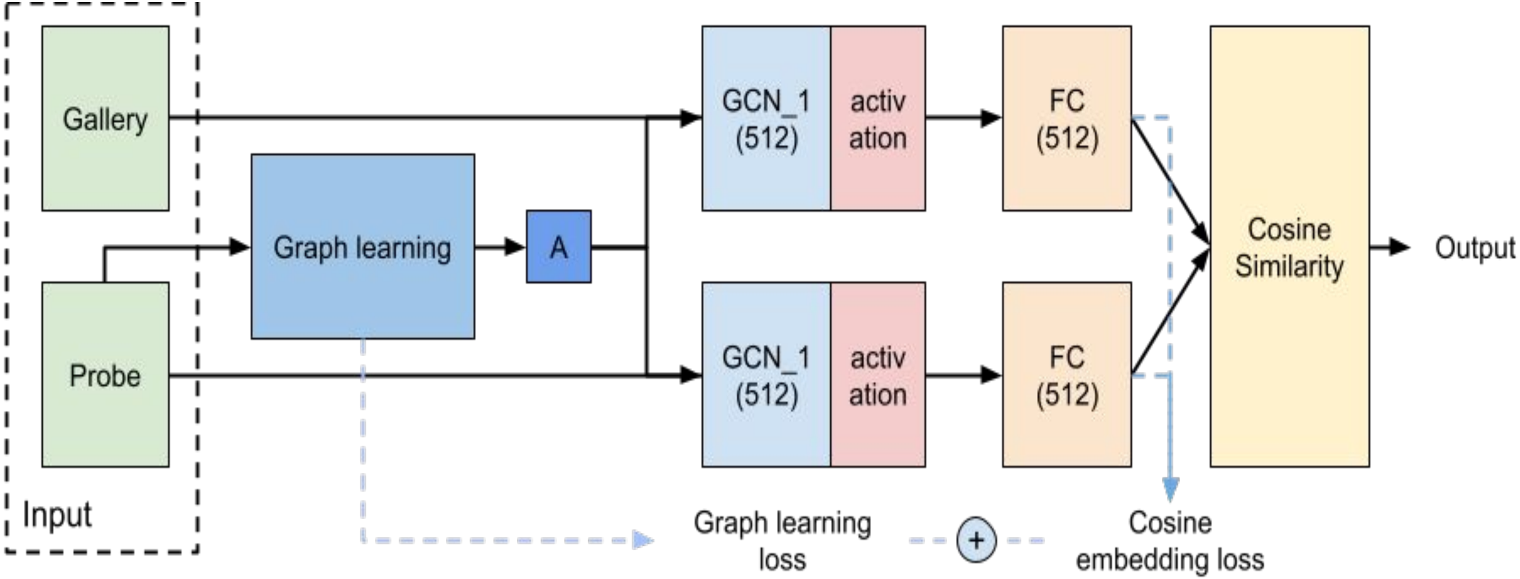
Graph convolution modified activation network



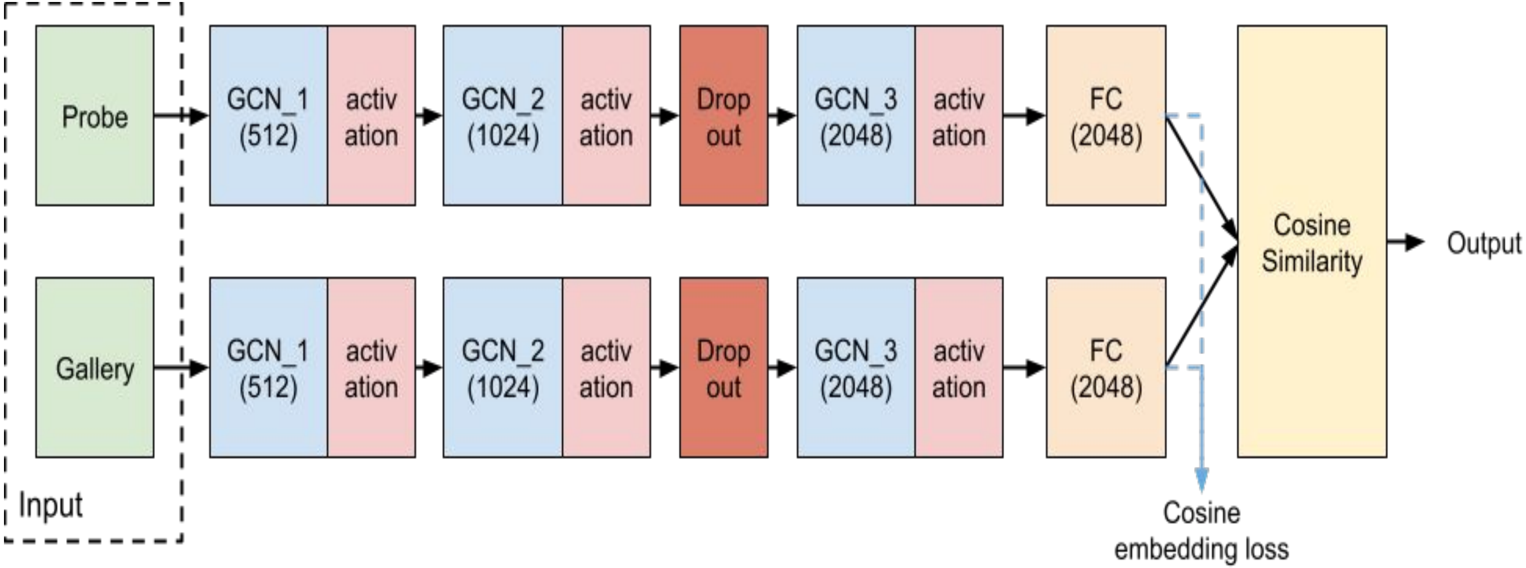
Graph attention with graph convolution network



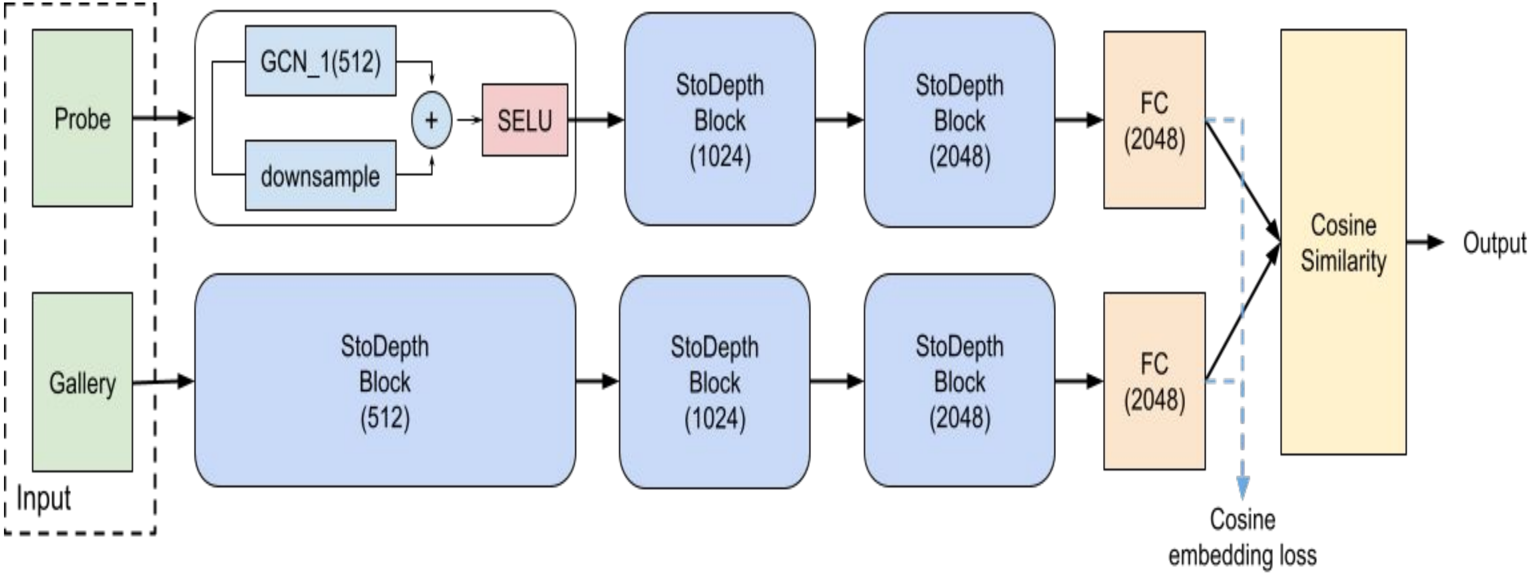
Graph learning combined graph convolution network



Dropout base on level hops



Stochastic Depth in Graph Convolution network



RESULTS ANALYSIS

Table 4.1: Performance evaluate Data processing base on Gcn base model with 3 layer and CUHK-SYSU dataset [1].

Data processing	mAP (%)	top-1 (%)	top-5 (%)
Original	80.30	82.00	91.79
Re-make	81.88	83.14	92.76

RESULTS ANALYSIS

Table 4.2: Performance evaluate our model Gcn base model on CUHK-SYSU dataset with gallery size of 100

Model	mAP (%)	top-1 (%)	top-5 (%)	top-10 (%)
Gcn 3 layer [1]	81.80	83.14	92.59	95.07
Gcn 2 layer [1]	81.88	82.66	92.86	95.03
Gcn 3 remove ReLU	82.57	83.90	93.10	**95.48**
Gcn 2 remove ReLU	82.63	83.86	93.52	95.38
Gcn 3 CeLU	82.62	84.00	93.31	95.34
Gcn 3 SeLU	82.80	84.21	93.28	95.31
Gcn 2 SeLU	**82.88**	**84.34**	93.24	95.28
GAT + GCN	81.45	82.34	92.76	94.34
GLCN	82.06	83.72	**93.66**	95.24

RESULTS ANALYSIS

Table 4.3: Performance evaluate our model with noisy training.

Regularization	mAP (%)	top-1 (%)	top-5 (%)	top-10 (%)
Gcn 3 SeLU	**82.80**	**84.21**	93.28	95.31
Gcn 3 Stochastic Depth	81.51	82.55	92.48	94.59
Gcn 3 Dropout	82.06	83.14	92.97	94.90

CONCLUSIONS

The technology landscape is continuously developing and changing. Demand for advanced technologies never stops increasing as people constantly tackle new challenges or make new discoveries. Besides that, the demand for identification technology will be increasing, as it is the basis for user support applications as well as collective management. To keep up with that trend, the optimization of the computational complexity of the RE-ID task is critical. It can be said that the application of the context has actually opened a path toward where we can make more accurate predictions at a low computational cost.

We have made two propositions in this thesis. First is the improvement of the data loading method described in Chapter 4. Second, studies were conducted to expand and improve the model's learning capabilities through the graphs' application. Along with the coupling from existing models, the proposed models include a GCN architecture based on selu rather than the conventional GCN architecture based on Relu stated in Chapter 3. This method gives better identification results than the traditional method mentioned in the author's research.

CONCLUSIONS

We discovered during the experiment that maintaining a definition of the directed adjacency matrix assists the model in learning. The result is superior in studies using adaptive models such as GAT or GLCN. This is due to the orientation supporting the model in learning effectively from the beginning, allowing it to progress further under the same time limit. Besides, the expansion of the nonlinear space from relu to selu has also shown a positive effect as the model can learn better.

In future work, we will develop our simultaneous multi-label recognition institute deeply with the desire to maximize the subject defined per frame. Then the use of more adaptive networks will be able to exert its effectiveness in learning complex contexts instead of the oriented contexts as in the experiment. On the other hand, there is the application of noise learning methods to improve model quality by breaking down the hops restriction. The limitation of Hops can only be solved when we can represent the relationships of the nodes referenced through an adjacency matrix or have to pass information from the head layers straight to the deeper layers. That is a direction that we rationally expect will improve the ability of the model to learn.

Thank you for your attention!
Feel free to ask any questions.

Contacts:
Duy Le Dinh, duyldhe130655@fpt.edu.vn
Student of Computer Science
FPT University
Hanoi, Viet Nam

Website
http://fpt.edu.vn/