



# Learning Graph Convolutional Network for Skeleton-based Human Action Recognition by Neural Searching

Nguyen Anh Tung

Student ID: 2020 – 25375

Seoul National University

# CONTENT



PROBLEM



METHOD



EVALUATION

A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines. The nodes are represented by small circles, some of which are highlighted with a double-circle outline. The lines are thin and gray, creating a mesh-like structure.

1.

# PROBLEM

Let's start with the problem we  
need to solve

# PROBLEM

- Human action recognition from Skeleton data fueled by the Graph Convolutional Network.
- Most GCN methods ignores implicit joint correlations.
- The need of replacing the fixed graph structure with dynamic one.
- Higher-order connections are not well involved



A decorative network diagram in the top-left corner, featuring a complex web of interconnected nodes and lines, with some nodes highlighted in grey and others in white.

# 2.

# METHOD

Let's continue with the proposed method

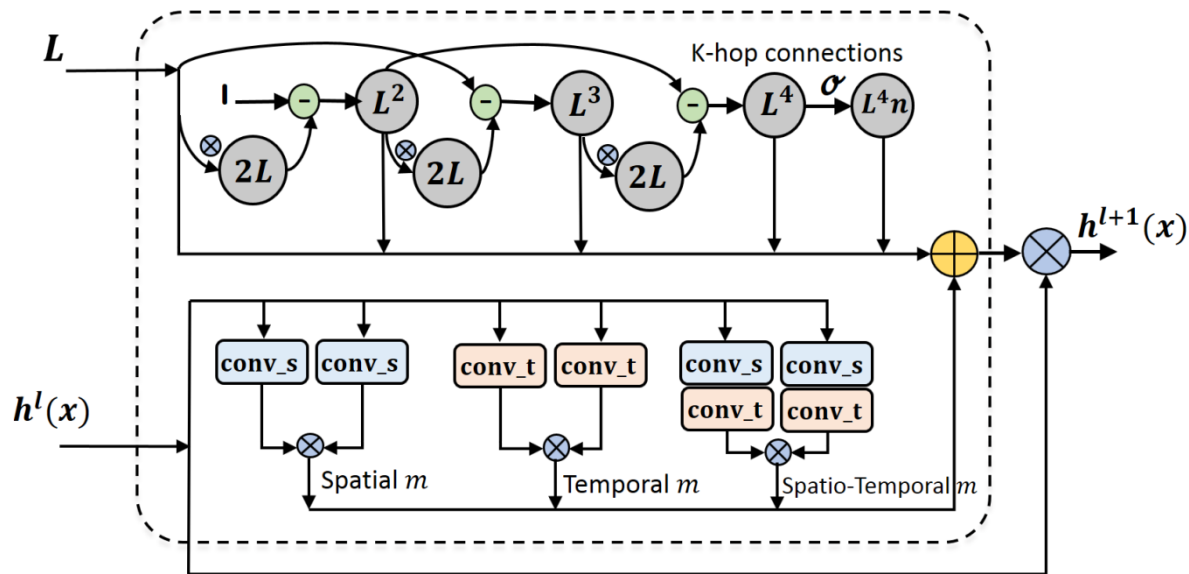
# METHOD

- ◎ Using Automatic Neural Architecture Search (NAS)
- ◎ Searching in a GCN space built with multiple graph function modules.
- ◎ Using Search strategy for both sampling and memory efficient.

# NAS



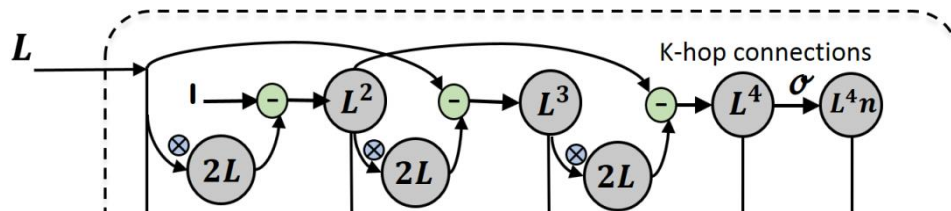
# METHOD



Chebyshev Polynomials:  $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$



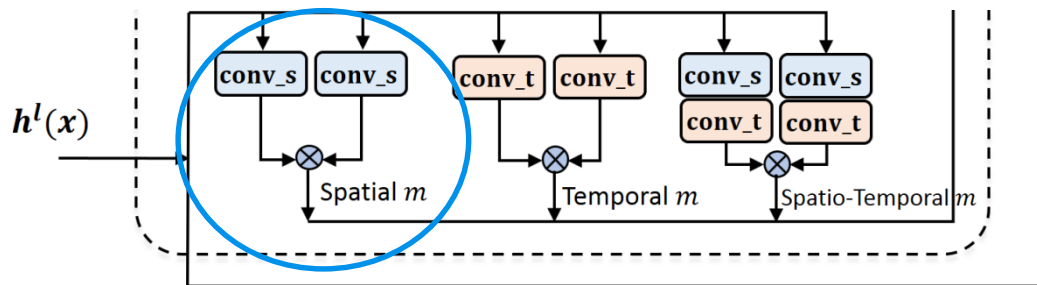
# METHOD



The Chebyshev polynomial functions let the network determine which order and polynomial components each layer prefers

$$\text{Chebyshev Polynomials: } T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$$

# METHOD



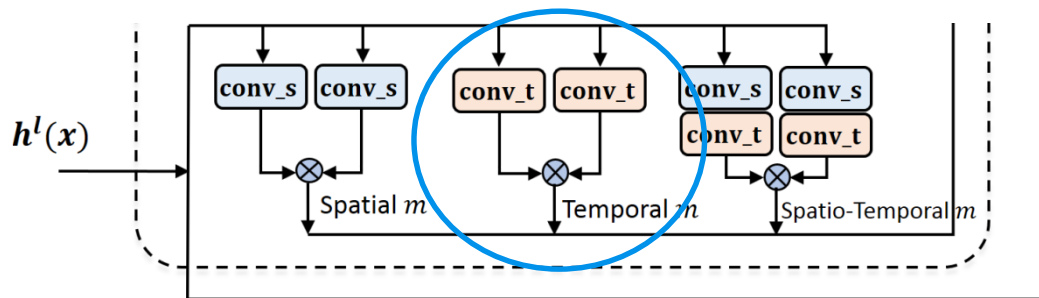
GCN  
SEARCH  
SPACE

Structure Representation Correlation as **Spatial m**:  
Compute based on the spatial node connections

Gaussian Function:

$$\forall i, j \in \mathcal{V}, A_D(i, j) = \frac{e^{\phi(h(x_i))} \otimes \psi(h(x_j))}{\sum_{j=1}^n e^{\phi(h(x_i))} \otimes \psi(h(x_j))}$$

# METHOD

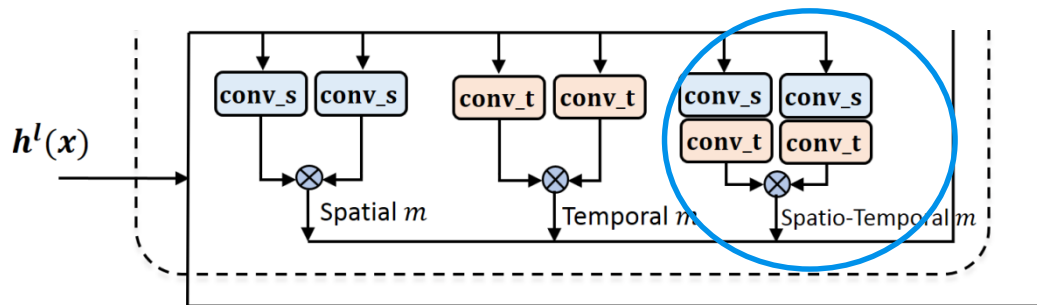


Temporal Representation Correlation as **Temporal m**:  
Extract the temporal information of each node before computing node correlations.

Guassian Function:

$$\forall i, j \in \mathcal{V}, A_D(i, j) = \frac{e^{\phi(h(x_i))} \otimes \psi(h(x_j))}{\sum_{j=1}^n e^{\phi(h(x_i))} \otimes \psi(h(x_j))}$$

# METHOD



GCN  
SEARCH  
SPACE

**Spatial-Temporal function:** combining the Spatial and Temporal Module

Gaussian Function:

$$\forall i, j \in \mathcal{V}, A_D(i, j) = \frac{e^{\phi(h(x_i))} \otimes \psi(h(x_j))}{\sum_{j=1}^n e^{\phi(h(x_i))} \otimes \psi(h(x_j))}$$

# SEARCH STRATEGY



A Venn diagram consisting of three overlapping circles with dashed blue borders. The left circle is labeled 'Cross-Entropy', the middle circle is labeled 'CEIM', and the right circle is labeled 'Importance-Mixing'. The circles overlap in pairs and in the center.

Cross-Entropy

CEIM

Importance  
-Mixing

# CEIM

## Three steps:

- ① **Sampling populations:** Modeling the architecture distribution with a Gaussian distribution.
- ② **Selecting populations:** Combining  $S_{\text{new}}$  with historical selected populations  $S_{\text{old}}$
- ③ **Updating:** Using new selected samples to update the architecture distribution.

( $S_{\text{new}}$ : the populations for CEIM)



3.

# EVALUATION

Let's end with the proposed  
method evaluation

# EVALUATION

Methods	Joint (%)	Bone (%)	Combine (%)
2S-AGCN	93.7	93.2	95.1
T	93.8	93.7	95.1
ST	94.0	93.8	95.2
T+Cheb	94.0	93.9	95.3
ST+Cheb	94.2	93.9	95.3
S+T+ST+Cheb	93.9	93.6	95.1
<b>NAS</b>	<b>94.6</b>	<b>94.7</b>	<b>95.7</b>

Performance Comparison on NTU RGB+D with CV evaluation



# EVALUATION

Methods	CS (%)	CV (%)
DPRL	83.5	89.8
SR-TSL	84.8	92.4
STGR-GCN	86.9	92.3
GR-GCN	87.5	94.3
AS-GCN	86.8	94.2
2S-AGCN	88.5	95.1
<b>NAS (Joint+Bone)</b>	<b>89.4</b>	<b>95.7</b>

Performance Comparison on NTU RGB+D with six current state-of-the-art methods

# EVALUATION

Methods	Top-1(%)	Top-5(%)
P-LSTM	16.4	25.8
ST-GCN	30.7	52.8
AS-GCN	34.8	56.5
2S-AGCN (Joint)	35.1	57.1
2S-AGCN (Bone)	33.3	55.7
2S-AGCN	36.1	58.7
<b>NAS (Joint)</b>	<b>35.5</b>	<b>57.9</b>
<b>NAS (Bone)</b>	<b>34.9</b>	<b>57.1</b>
<b>NAS (Joint+Bone)</b>	<b>37.1</b>	<b>60.1</b>

Performance Comparison on Kinetics with eight current state-of-the-art methods

# SUMMARY

- ◎ Determine the graph convolution architecture with NAS for skeleton-based action recognition.
- ◎ Provide multiple dynamic graph substructures on the basis of various spatial-temporal graph modules
- ◎ Build higher-order connections with Chebyshev polynomial.
- ◎ Use CEIM for improving the search efficiency

The background of the slide is a light gray network pattern. It consists of numerous small circles, some of which are white with a gray outline, and others are solid gray. These circles are interconnected by thin, light gray lines, creating a complex, web-like structure that fills the entire background.

**THANKS FOR  
LISTENING!**