# Ad-HGformer: An Adaptive HyperGraph Transformer for Skeletal Action Recognition -:ADDED MATERIAL:-

Anonymous ECCV 2024 Submission

Paper ID #10985

**Table 1:** The performance of various SOTA methods with the proposed **Ad**aptive **Hyp**ergraph **Dec**oder (Ad HypDec). * indicates the performance in our settings.

| **Dataset** | NTU-60 | | NTU-120 | | NW-UCLA |
|---|---|---|---|---|---|
| **Setting** | **X-Sub** | **X-View** | **X-Sub** | **X-View** | |
| DST-HCN | 90.7 | 96.0 | 86.0 | 87.9 | - |
| DST-HCN* | 90.7 | 95.9 | 85.9 | 87.9 | 94.9 |
| **DST-HCN*+Ad HypDec** | **91.3** | **96.4** | **86.5** | **88.3** | **95.1** |
| Selective-HCN | 90.8 | 96.6 | - | - | - |
| Selective-HCN* | 90.7 | 96.6 | 86.3 | 88.1 | 95.0 |
| **Selective-HCN*+Ad HypDec** | **91.4** | **97.0** | **86.7** | **88.3** | **95.3** |
| Hyperformer | 92.9 | 96.5 | 89.9 | 91.3 | 96.9 |
| Hyperformer* | 92.9 | 96.4 | 89.9 | 91.3 | 96.8 |
| **Hyperformer*+Ad HypDec** | **93.3** | **97.0** | **90.2** | **91.6** | **97.2** |
| 3Mformer | 94.8 | 98.7 | 92.0 | 93.8 | 97.8 |
| 3Mformer* | 94.8 | 98.6 | 91.9 | 93.8 | 97.8 |
| **3Mformer*+Ad HypDec** | **95.2** | **98.9** | **92.2** | **94.1** | **98.1** |

**Table 2:** Impact of different units of the proposed Ad-HGformer.

| **Model** | **Adaptive Hypergraph** | **Temporal Convolution** | **Temporal Attention** | **Accuracy in (%)** |
|---|---|---|---|---|
| Base Model + | | ✓ | ✓ | 93.07 |
| | ✓ | | | 93.30 |
| | ✓ | ✓ | | 93.38 |
| | ✓ | | ✓ | 93.41 |
| | ✓ | ✓ | ✓ | 93.50 |

**Table 3:** Impact of scaling factor $\alpha$ in the performance.

| $\alpha$ | **0.1** | **0.2** | **0.3** | **0.4** | **0.6** |
|---|---|---|---|---|---|
| Accuracy (%) | 93.14 | 93.50 | 93.42 | 93.31 | 93.31 |

**Table 4:** The significance of various modules in Ad-HGformer compared to baseline Hyperformer in terms of class-wise accuracy (%). ADG: Adaptive Hypergraph generator, RL: Reconstruction Loss, THA: Temporal Hypergraph Attention, CHA: CHannel Attention. The enhanced and reduced class-wise accuracy are given by ↑ and ↑ respectively.

| **Methods** | **Baseline** | **AHG** | **Ad+RL** | **Ad+RL+THA** | **Ad+RL+THA+CHA** |
|---|---|---|---|---|---|
| **Params(M)** / **Action Labels** | 2.60 | 2.75 | 2.95 | 3.10 | 3.20 |
| drink water | 86.57 | 87.15 (0.61↑) | 87.52 | 88.25 | 88.95 |
| eat meal/snack | 77.76 | 78.55(0.79↑) | 77.78 | 80.85 | 81.35 |
| brushing teeth | 90.55 | 91.15(0.60↑) | 91.20 | 91.54 | 91.86 |
| brushing hair | 91.31 | 92.57(1.26↑) | 93.95 | 94.12 | 94.26 |
| drop | 92.53 | 93.59(1.06↑) | 94.35 | 94.56 | 94.78 |
| pickup | 97.25 | 97.89(0.64↑) | 98.25 | 98.45 | 98.45 |
| throw | 93.35 | 93.87(0.52↑) | 93.95 | 94.10 | 94.25 |
| sitting down | 98.75 | 98.90(0.15↑) | 99.10 | 99.27 | 99.27 |
| standing up (from sitting position | 98.75 | 98.90(0.15↑) | 98.90 | 99.15 | 99.27 |
| clapping | 84.86 | 86.45(1.59↑) | 86.85 | 87.00 | 87.10 |
| reading | 60.66 | 63.46(2.80↑) | 63.92 | 64.02 | 64.18 |
| writing | 67.50 | 71.56(4.06↑) | 71.85 | 72.00 | 72.32 |
| tear up paper | 95.36 | 95.28(0.08↓) | 95.36 | 95.46 | 95.57 |
| wear jacket | 98.45 | 98.36(0.09↓) | 98.45 | 98.55 | 98.55 |
| take off jacket | 98.82 | 98.82(0.00↑) | 99.02 | 99.18 | 99.18 |
| wear a shoe | 65.75 | 81.66(15.91↑) | 83.74 | 84.16 | 84.78 |
| take off a shoe | 82.75 | 83.75(0.00↑) | 83.75 | 84.00 | 84.00 |
| wear on glasses | 94.12 | 94.45(0.33↑) | 94.67 | 94.95 | 95.15 |
| take off glasses | 95.33 | 95.12(0.21↓) | 95.56 | 95.58 | 95.62 |
| put on a hat/cap | 98.16 | 98.65(0.49↑) | 98.65 | 98.75 | 98.75 |
| take off a hat/cap | 98.95 | 98.83(0.12↓) | 98.89 | 98.95 | 98.95 |
| cheer up | 93.70 | 94.55(0.85↑) | 94.80 | 94.89 | 94.89 |
| hand waving | 94.00 | 94.65(0.65↑) | 95.10 | 95.25 | 95.35 |
| kicking something | 96.63 | 97.57(0.94↑) | 97.68 | 97.85 | 97.85 |
| reach into pocket | 86.29 | 86.29(0.00↑) | 86.37 | 86.37 | 86.37 |
| hopping (one foot jumping) | 98.81 | 98.91(0.10↑) | 98.91 | 98.91 | 98.91 |
| jump up | 99.25 | 99.25(0.00↑) | 99.25 | 99.25 | 99.25 |
| make a phone call/answer phone | 92.00 | 92.20(0.20↑) | 92.35 | 92.55 | 92.55 |
| playing with phone/tablet | 77.67 | 81.25(3.58↑) | 81.38 | 81.47 | 82.56 |
| typing on a keyboard | 72.45 | 75.82(3.37↑) | 77.27 | 77.66 | 77.86 |
| pointing to something with finger | 81.75 | 86.89(5.14↑) | 87.95 | 87.34 | 87.49 |
| taking a selfie | 94.29 | 94.45(0.16↑) | 94.66 | 94.66 | 94.75 |
| check time (from watch) | 93.19 | 93.56(0.37↑) | 93.85 | 94.00 | 94.25 |
| rub two hands together | 91.65 | 91.85(0.20↑) | 92.10 | 92.25 | 92.36 |
| nod head/bow | 98.85 | 99.00(0.15↑) | 99.18 | 99.18 | 99.18 |
| shake head | 96.35 | 96.25(0.10↓) | 96.35 | 96.35 | 96.35 |
| wipe face | 87.26 | 89.35(2.09↑) | 89.89 | 90.05 | 90.17 |
| salute | 95.25 | 95.25(0.00↑) | 95.47 | 95.55 | 95.55 |
| put the palms together | 98.86 | 98.24(0.62↓) | 98.46 | 98.65 | 98.86 |
| cross hands in front (say stop) | 97.46 | 98.00(0.54↑) | 98.15 | 98.15 | 98.15 |
| sneeze/cough | 79.85 | 84.15(4.30↑) | 84.45 | 84.95 | 85.69 |
| staggering | 99.48 | 99.28(0.20↓) | 99.48 | 99.48 | 99.48 |
| falling | 99.52 | 99.52(0.00↑) | 99.52 | 99.52 | 99.52 |
| touch head (headache) | 85.85 | 87.63(1.78↑) | 88.00 | 88.55 | 88.87 |
| touch chest (stomachache/heart pain) | 95.55 | 96.00(0.45↑) | 96.25 | 96.55 | 96.78 |
| touch back (backache) | 96.36 | 96.36(0.00↑) | 96.48 | 96.48 | 96.48 |
| touch neck (neckache) | 90.44 | 92.87(2.43↑) | 92.97 | 93.25 | 93.44 |
| nausea or vomiting condition | 86.95 | 87.05(0.10↑) | 87.22 | 87.22 | 87.22 |
| use a fan/feeling warm | 91.20 | 93.67(2.47↑) | 93.89 | 94.00 | 94.26 |
| punching/slapping other person | 94.10 | 94.64(0.54↑) | 94.72 | 94.72 | 94.89 |
| kicking other person | 96.24 | 96.53(0.29↑) | 96.53 | 96.68 | 96.78 |
| pushing other person | 99.15 | 99.25(0.10↑) | 99.25 | 99.25 | 99.25 |
| pat on back of other person | 94.58 | 95.36(0.78↑) | 95.36 | 95.68 | 95.88 |
| point finger at the other person | 93.67 | 94.57(0.90↑) | 95.00 | 95.45 | 95.68 |
| hugging other person | 99.55 | 99.55(0.00↑) | 99.55 | 99.55 | 99.55 |
| giving something to other person | 96.42 | 96.72(0.30↑) | 96.72 | 96.83 | 96.83 |
| touch other person's+pocket | 98.21 | 98.55(0.34↑) | 98.62 | 98.75 | 98.91 |
| handshaking | 98.14 | 98.28(0.14↑) | 98.28 | 98.36 | 98.36 |
| walking towards each other | 99.56 | 99.66(0.10↑) | 99.66 | 99.66 | 99.66 |
| walking apart from each other | 98.22 | 98.63(0.41↑) | 98.72 | 98.81 | 98.89 |
| **average** | 92.90 | 93.30(0.40↑) | 93.39 | 93.45 | 93.50 |

**Message to Reviewer#2** For any graph application, the higher-order relation of each node with the other nodes and links plays a major role. Therefore, joint-joint self-attention and joint-bone cross-attention. Some groups of nodes are highly co-related; therefore, hyperedge and joint-hyperedge self-attention. First, we think of hyperedge-hyperedge self-attention for "how one hyperedge related to another hyperedge." But the parameter increases vastly with the increase in performance. So, instead, we proposed temporal hyperedge attention applied to alternative blocks that effectively enhance the accuracy at the expense of fewer parameters. The idea of adaptive hyperedge comes as highly co-related nodes in hyperedge must be varied from one action class to another. The presence of the decoder makes the parameter learnable in an unsupervised manner and safeguards the clustering from the "curse of dimensionality."