|  |  |
| --- | --- |
| MODEL | METEOR |
| LSTM (CNN with RNN framework, mean pooling)  LSTM( with max pooling) | 26.9  28.9 |
| S2VT (RGB)  S2VT (Optical flow)  LSTM GAN (without attention)  LSTM (with attention) | 29.2  29.8  29.8  30.6 |

MSVD Dataset

|  |  |
| --- | --- |
| Using Language Model  Using Language model + Discriminator model | 27.4  28 |

|  |  |
| --- | --- |
| Using Language Model | 26.4 |
| Using Language Model+ Discrimator | 27.3 |

MSR-VTT Dataset

|  |  |  |
| --- | --- | --- |
|  | MSVD | MSR-VTT |
| Videos  Sentences  Avg length | 1970  80,827  10.2s | 68,337  68,375  3.9s |