Supervised learning:

Training: Generating caption sequence $Y = \{y_1, \ldots, y_T\}, y_t \in \mathbb{D}$ given an image $I$ by cross entropy (XE) loss: $C(Y) = \sum_{t=1}^{T} \log p(y_t | y_1, \ldots, y_{t-1}).$

Testing: The trained sequence generative model is evaluated by computing the task-specific score $R(Y, Y)$ (e.g., BLEU, CIDEr) on the test set, where $Y$ is the predicted caption sequence.

Challenges:

- (i) Training: maximize the likelihood of each ground-truth word given the previous ground-truth words and the image, termed Teacher-Forcing. Testing: the model uses the previously generated words from the model distribution to predict the next word. This exposure bias can result in error accumulation in sentence generation during test time, since the model has never been exposed to its own predictions.

- (ii) The training supervision metric, such as the widely used cross entropy loss, is different from the evaluation metrics at test time. The evaluation metrics are non-differentiable.

Related work:

Policy gradient [4] adds an additional FC layer on top of the RNN output to predict state value function. However, it treats $\text{state}$ as the RNN output while we treat the $\text{state}$ as the RNN input (given image and the taken actions), so that we can build an independent value network rather than a shared RNN cell between actor and critic. 

Methodology:

```
Methodology:

Actor-Critic Sequence Training for Image Captioning
Li Zhang    Flood Sung    Feng Liu    Tao Xiang    Shaogang Gong    Yongxin Yang    Timothy M. Hospedales
(david.li.zhang, feng.liu, t.xiang, s.gong, yongxin.yang)@qmul.ac.uk    floodsung@gmail.com
The University of Edinburgh    t.hospedales@ed.ac.uk

Supervised learning:

Training: Generating caption sequence $Y = \{y_1, \ldots, y_T\}, y_t \in \mathbb{D}$ given an image $I$ by cross entropy (XE) loss: $C(Y) = \sum_{t=1}^{T} \log p(y_t | y_1, \ldots, y_{t-1}).$

Testing: The trained sequence generative model is evaluated by computing the task-specific score $R(Y, Y)$ (e.g., BLEU, CIDEr) on the test set, where $Y$ is the predicted caption sequence.

Challenges:

- (i) Training: maximize the likelihood of each ground-truth word given the previous ground-truth words and the image, termed Teacher-Forcing. Testing: the model uses the previously generated words from the model distribution to predict the next word. This exposure bias can result in error accumulation in sentence generation during test time, since the model has never been exposed to its own predictions.

- (ii) The training supervision metric, such as the widely used cross entropy loss, is different from the evaluation metrics at test time. The evaluation metrics are non-differentiable.

Related work:

Policy gradient [4] adds an additional FC layer on top of the RNN output to predict state value function. However, it treats $\text{state}$ as the RNN output while we treat the $\text{state}$ as the RNN input (given image and the taken actions), so that we can build an independent value network rather than a shared RNN cell between actor and critic.

Methodology:

Actor-Critic Sequence Training for Image Captioning
Li Zhang    Flood Sung    Feng Liu    Tao Xiang    Shaogang Gong    Yongxin Yang    Timothy M. Hospedales
(david.li.zhang, feng.liu, t.xiang, s.gong, yongxin.yang)@qmul.ac.uk    floodsung@gmail.com
The University of Edinburgh    t.hospedales@ed.ac.uk

Supervised learning:

Training: Generating caption sequence $Y = \{y_1, \ldots, y_T\}, y_t \in \mathbb{D}$ given an image $I$ by cross entropy (XE) loss: $C(Y) = \sum_{t=1}^{T} \log p(y_t | y_1, \ldots, y_{t-1}).$

Testing: The trained sequence generative model is evaluated by computing the task-specific score $R(Y, Y)$ (e.g., BLEU, CIDEr) on the test set, where $Y$ is the predicted caption sequence.

Challenges:

- (i) Training: maximize the likelihood of each ground-truth word given the previous ground-truth words and the image, termed Teacher-Forcing. Testing: the model uses the previously generated words from the model distribution to predict the next word. This exposure bias can result in error accumulation in sentence generation during test time, since the model has never been exposed to its own predictions.

- (ii) The training supervision metric, such as the widely used cross entropy loss, is different from the evaluation metrics at test time. The evaluation metrics are non-differentiable.

Related work:

Policy gradient [4] adds an additional FC layer on top of the RNN output to predict state value function. However, it treats $\text{state}$ as the RNN output while we treat the $\text{state}$ as the RNN input (given image and the taken actions), so that we can build an independent value network rather than a shared RNN cell between actor and critic.

Methodology:

Actor-Critic Sequence Training for Image Captioning
Li Zhang    Flood Sung    Feng Liu    Tao Xiang    Shaogang Gong    Yongxin Yang    Timothy M. Hospedales
(david.li.zhang, feng.liu, t.xiang, s.gong, yongxin.yang)@qmul.ac.uk    floodsung@gmail.com
The University of Edinburgh    t.hospedales@ed.ac.uk

Supervised learning:

Training: Generating caption sequence $Y = \{y_1, \ldots, y_T\}, y_t \in \mathbb{D}$ given an image $I$ by cross entropy (XE) loss: $C(Y) = \sum_{t=1}^{T} \log p(y_t | y_1, \ldots, y_{t-1}).$

Testing: The trained sequence generative model is evaluated by computing the task-specific score $R(Y, Y)$ (e.g., BLEU, CIDEr) on the test set, where $Y$ is the predicted caption sequence.

Challenges:

- (i) Training: maximize the likelihood of each ground-truth word given the previous ground-truth words and the image, termed Teacher-Forcing. Testing: the model uses the previously generated words from the model distribution to predict the next word. This exposure bias can result in error accumulation in sentence generation during test time, since the model has never been exposed to its own predictions.

- (ii) The training supervision metric, such as the widely used cross entropy loss, is different from the evaluation metrics at test time. The evaluation metrics are non-differentiable.

Related work:

Policy gradient [4] adds an additional FC layer on top of the RNN output to predict state value function. However, it treats $\text{state}$ as the RNN output while we treat the $\text{state}$ as the RNN input (given image and the taken actions), so that we can build an independent value network rather than a shared RNN cell between actor and critic.

Methodology:

Actor-Critic Sequence Training for Image Captioning
Li Zhang    Flood Sung    Feng Liu    Tao Xiang    Shaogang Gong    Yongxin Yang    Timothy M. Hospedales
(david.li.zhang, feng.liu, t.xiang, s.gong, yongxin.yang)@qmul.ac.uk    floodsung@gmail.com
The University of Edinburgh    t.hospedales@ed.ac.uk

Supervised learning:

Training: Generating caption sequence $Y = \{y_1, \ldots, y_T\}, y_t \in \mathbb{D}$ given an image $I$ by cross entropy (XE) loss: $C(Y) = \sum_{t=1}^{T} \log p(y_t | y_1, \ldots, y_{t-1}).$

Testing: The trained sequence generative model is evaluated by computing the task-specific score $R(Y, Y)$ (e.g., BLEU, CIDEr) on the test set, where $Y$ is the predicted caption sequence.

Challenges:

- (i) Training: maximize the likelihood of each ground-truth word given the previous ground-truth words and the image, termed Teacher-Forcing. Testing: the model uses the previously generated words from the model distribution to predict the next word. This exposure bias can result in error accumulation in sentence generation during test time, since the model has never been exposed to its own predictions.

- (ii) The training supervision metric, such as the widely used cross entropy loss, is different from the evaluation metrics at test time. The evaluation metrics are non-differentiable.

Related work:

Policy gradient [4] adds an additional FC layer on top of the RNN output to predict state value function. However, it treats $\text{state}$ as the RNN output while we treat the $\text{state}$ as the RNN input (given image and the taken actions), so that we can build an independent value network rather than a shared RNN cell between actor and critic.

Methodology:

Actor-Critic Sequence Training for Image Captioning
Li Zhang    Flood Sung    Feng Liu    Tao Xiang    Shaogang Gong    Yongxin Yang    Timothy M. Hospedales
(david.li.zhang, feng.liu, t.xiang, s.gong, yongxin.yang)@qmul.ac.uk    floodsung@gmail.com
The University of Edinburgh    t.hospedales@ed.ac.uk

Supervised learning:

Training: Generating caption sequence $Y = \{y_1, \ldots, y_T\}, y_t \in \mathbb{D}$ given an image $I$ by cross entropy (XE) loss: $C(Y) = \sum_{t=1}^{T} \log p(y_t | y_1, \ldots, y_{t-1}).$

Testing: The trained sequence generative model is evaluated by computing the task-specific score $R(Y, Y)$ (e.g., BLEU, CIDEr) on the test set, where $Y$ is the predicted caption sequence.

Challenges:

- (i) Training: maximize the likelihood of each ground-truth word given the previous ground-truth words and the image, termed Teacher-Forcing. Testing: the model uses the previously generated words from the model distribution to predict the next word. This exposure bias can result in error accumulation in sentence generation during test time, since the model has never been exposed to its own predictions.

- (ii) The training supervision metric, such as the widely used cross entropy loss, is different from the evaluation metrics at test time. The evaluation metrics are non-differentiable.

Related work:

Policy gradient [4] adds an additional FC layer on top of the RNN output to predict state value function. However, it treats $\text{state}$ as the RNN output while we treat the $\text{state}$ as the RNN input (given image and the taken actions), so that we can build an independent value network rather than a shared RNN cell between actor and critic.