

عنوان بندی تصویر

کارگاه یادگیری عمیق با پایتون

سید ناصر رضوی www.snrazavi.ir

۱۳۹۷

پردازش زبان طبیعی



یک میز بلند که دور آن صندلی چوبی چیده شده و روی آن یک گلدان قرار دارد.

- دسته‌بندی متون
- مدل‌سازی زبان
- عنوان‌بندی تصاویر
- ترجمه ماشینی

فهرست مطالب

- عنوان‌بندی تصویر
- دادگان: میکروسافت کوکو
- معماری رمزگزار-رمزگشا
- معیارهای ارزیابی
- بهبود نتایج

معرفی مسئله

۴

□ عنوان‌بندی تصویر.

□ ورودی: یک تصویر

□ خروجی: یک توصیف متنی از تصویر داده شده



دو پسر بچه در حال انجام بازی هاکی

□ مجموعه داده مایکروسافت کوکو.

COCO 2015 Image Captioning Task

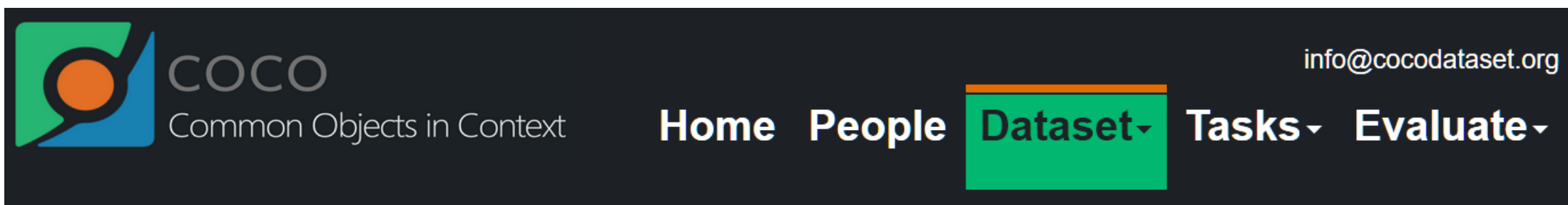


The man at bat readies to swing at the pitch while the umpire looks on.



A large bus sitting next to a very tall building.

مجموعه داده مایکروسافت کوکو. □



The screenshot shows the top navigation bar of the COCO dataset website. On the left is the COCO logo (a green square with a magnifying glass) and the text "COCO Common Objects in Context". To the right of the logo are navigation links: "Home", "People", "Dataset" (highlighted with a green background), "Tasks", and "Evaluate". The email address "info@cocodataset.org" is displayed in the top right corner.

Tools

[COCO API](#)

Images

- [2014 Train images \[83K/13GB\]](#)
- [2014 Val images \[41K/6GB\]](#)
- [2014 Test images \[41K/6GB\]](#)
- [2015 Test images \[81K/12GB\]](#)
- [2017 Train images \[118K/18GB\]](#)
- [2017 Val images \[5K/1GB\]](#)
- [2017 Test images \[41K/6GB\]](#)
- [2017 Unlabeled images \[123K/19GB\]](#)

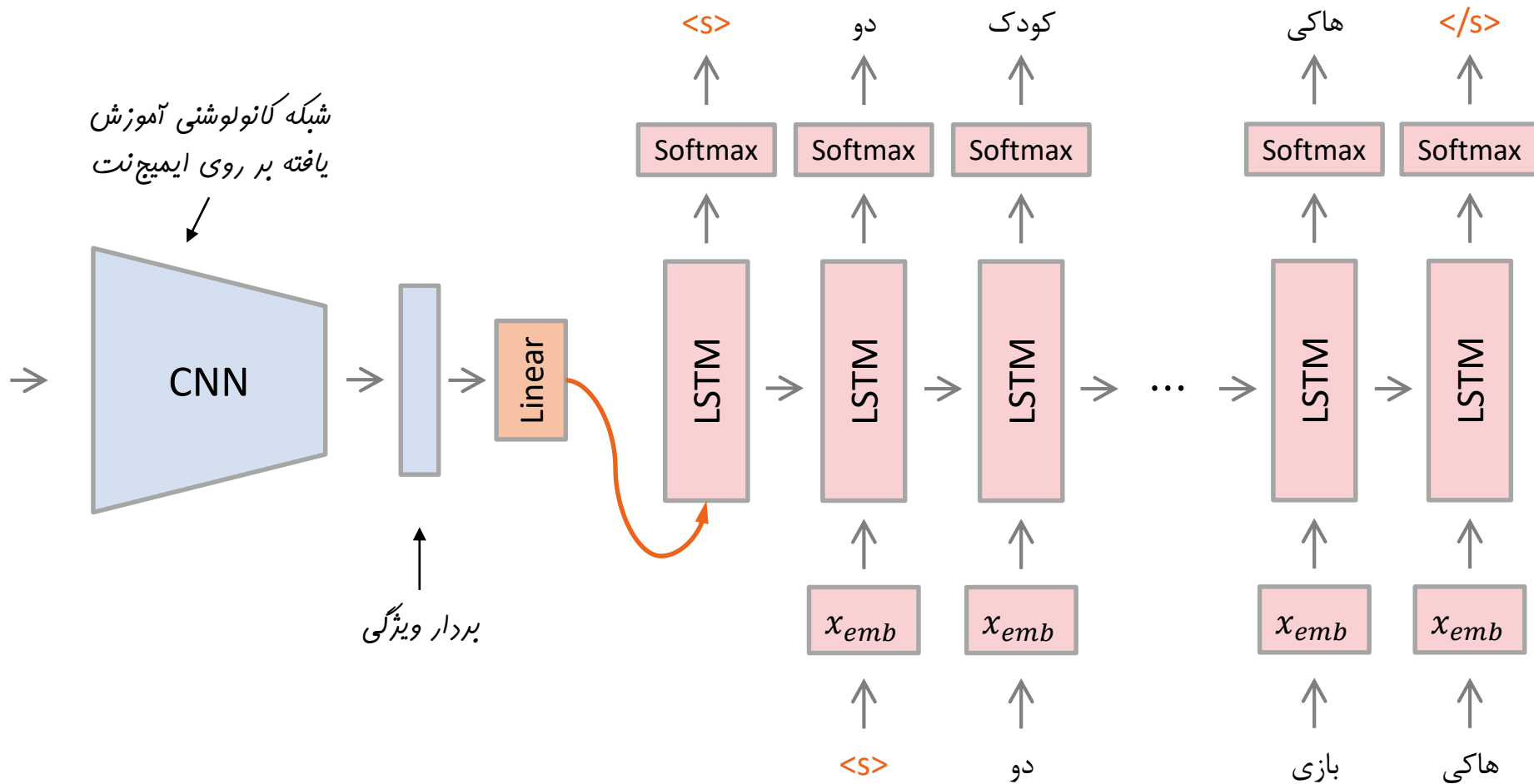
Annotations

- [2014 Train/Val annotations \[241MB\]](#)
- [2014 Testing Image info \[1MB\]](#)
- [2015 Testing Image info \[2MB\]](#)
- [2017 Train/Val annotations \[241MB\]](#)
- [2017 Stuff Train/Val annotations \[401MB\]](#)
- [2017 Testing Image info \[1MB\]](#)
- [2017 Unlabeled Image info \[4MB\]](#)

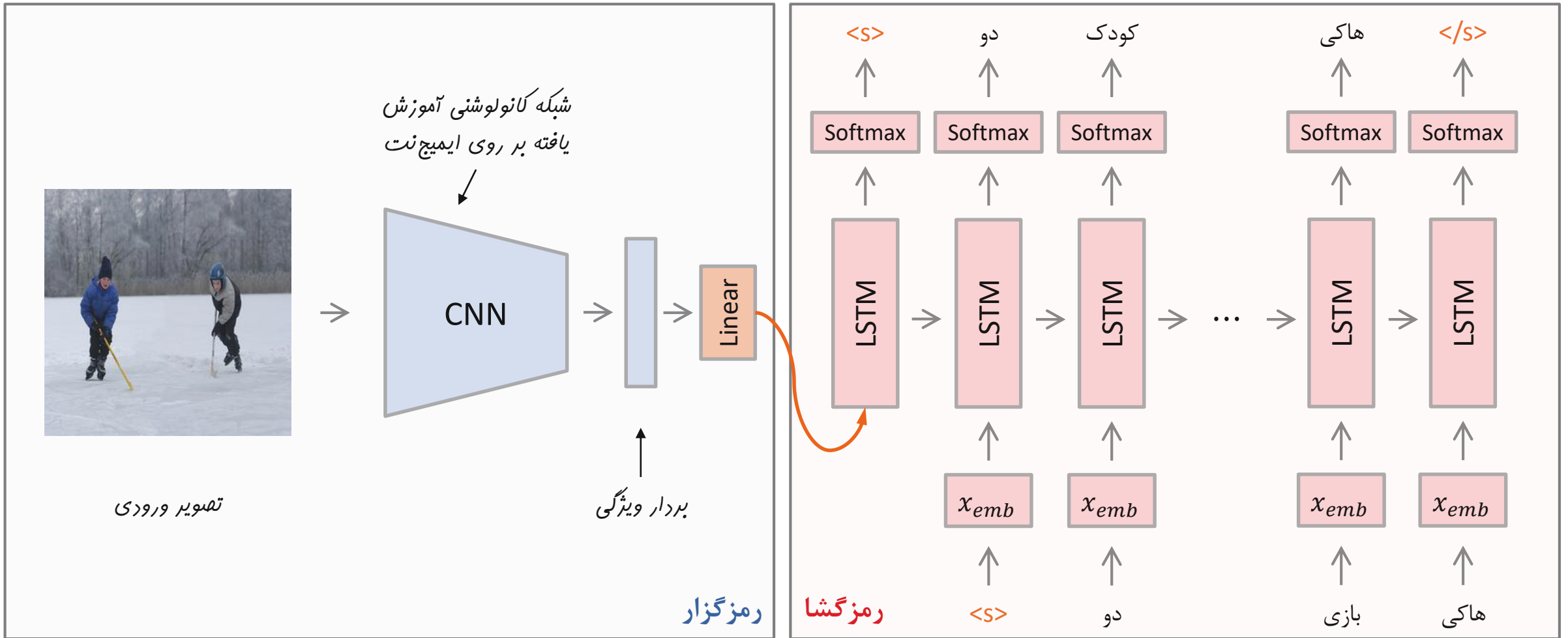


تصویر ورودی

شبکه کانولوشنی آموزش یافته بر روی ایمپجنت



معماری رمزگزار-رمزگشا




```
# create cocoEval object by taking coco and cocoRes
cocoEval = COCOEvalCap(coco, cocoRes)

# evaluate on a subset of images by setting
# cocoEval.params['image_id'] = cocoRes.getImgIds()
# please remove this line when evaluating the full validation set
cocoEval.params['image_id'] = cocoRes.getImgIds()

# evaluate results
# SPICE will take a few minutes the first time, but speeds up due to caching
cocoEval.evaluate()

tokenization...
setting up scorers...
computing Bleu score...
{'reflen': 9855, 'guess': [9893, 8893, 7893, 6893], 'testlen': 9893, 'correct': [5732, 2510, 1043, 423]}
ratio: 1.00385591071
Bleu_1: 0.579
Bleu_2: 0.404
Bleu_3: 0.279
Bleu_4: 0.191
computing METEOR score...
METEOR: 0.195
computing Rouge score...
ROUGE_L: 0.396
computing CIDEr score...
CIDEr: 0.600
computing SPICE score...
SPICE: 0.133
```



ground truth captions

- A train traveling down tracks next to lights .
- A blue and silver train next to train station and trees .
- A blue train is next to a sidewalk on the rails .
- A passenger train pulls into a train station .
- A train coming down the tracks arriving at a station .

generated caption

train traveling down a track in front of a road

بهبود نتایج

- دادگان.
- استفاده از داده‌های بیشتر
- رمزگزار.
- تنظیم دقیق لایه‌های آخر در رمزگزار
- رمزگشا.
- استفاده از بردارهای واژگان آموزش یافته بر روی یک پیکره بسیار بزرگ.
- استفاده از شبکه برگشتی دوسویه
- افزایش عمق شبکه برگشتی
- استفاده از جستجوی دسته‌ای به جای جستجوی حریصانه برای تولید عنوان‌ها
- استفاده از سازوکار توجه