

**JET
BRAINS**



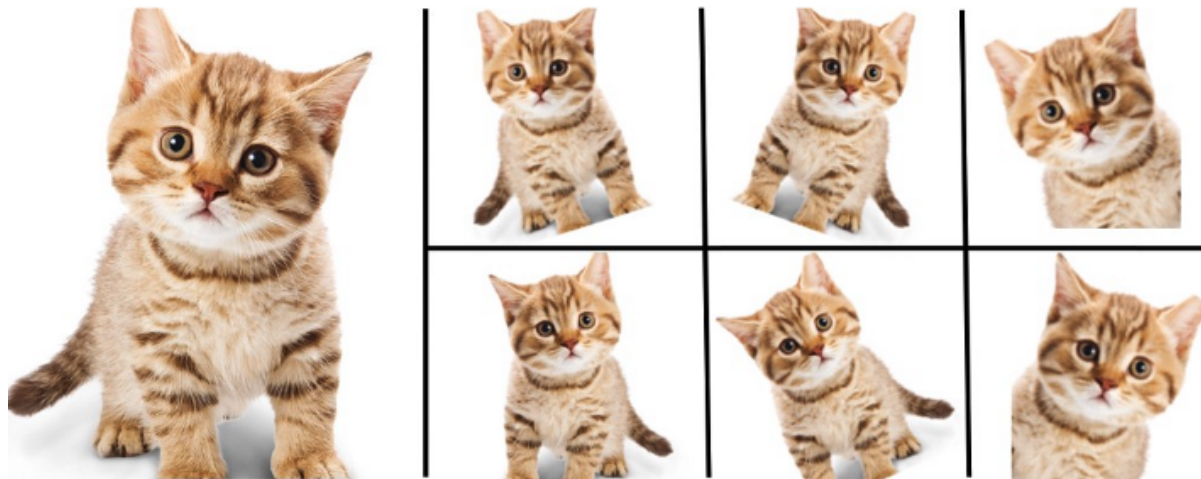
Text Augmentation



September, 2019
Moscow

Data augmentation in Computer Vision

- Flipping
- Rotation
- Clipping
- ...



Enlarge your Dataset

Data augmentation in NLP

Trivial:

- Change random words
- Forward-backward translation into other languages
- ...

```

I have to say, Seventeen & Missing is much better than I expected.
1 The
perception I took from the previews was that it would be just
humdrum
but I was pleasantly surprised with this impressive mystery.<br /
><br />Dedee Pfeiffer is Emilie, a mom who
insists her daughter, Lori (Tegan Moss), not attend a so-called
graduation
party one weeknight, but Lori ignores her mother's wishes and tak
es off for the party anyway. When Lori
does not come home, Emilie knows
something is wrong and she begins to have visions
of her daughter and the events that led to her disappearance.<br
/><br />Seventeen & Missing is better than so many other TV
movies
of this type, as it is not so predictable. Pfeiffer is the reason
to
see
this movie, and most of it comes off as believable. This LMN Orig
inal Movie premiered last night. 10/10

```

```

I have to say, Seventeen & Missing is much better than I expected.
1 The
image I attract from the preliminary study
was that it would be just lifeless
but I was pleasantly surprised with this impressive mystery.<br /
><br />Dedee Pfeiffer is Emilie, a mom who
swear her offspring, Lori (Tegan Moss), not appear a so-called
convocation
party one weeknight, but Lori ignores her mother's wishes and tak
es off for the party anyway. When Lori
make not come home, Emilie on top of
something is wrong and she begins to have eyes
of her daughter and the events that led to her disappearance.<br
/><br />Seventeen & Missing is better than so many other TV
cinematics
of this type, as it is not so predictable. Pfeiffer is the reason
to
behold
this movie, and most of it comes off as believable. This LMN Orig
inal Movie premiered last night. 10/10

```

Challenges

Effect of data augmentation can be measured on text classification tasks, either academic or from competition platforms.

- Formal languages
- Domain-specific texts
- Different impact for different classification models
- Impact on performance on imbalanced datasets
- Impact on model robustness

Results

—

- Develop a library for text augmentation
- Possible extension into topic of adversarial examples and adversarial robustness

**Thank you
for your attention**
