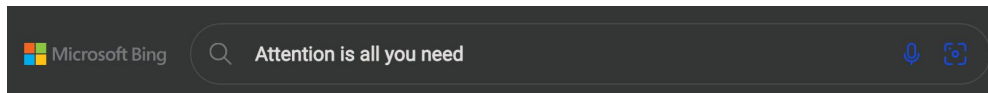


Reference Management

Papers for Research (Raw-Version)

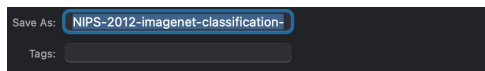
1. Search Paper



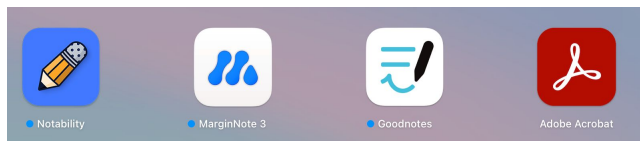
2. Download PDF



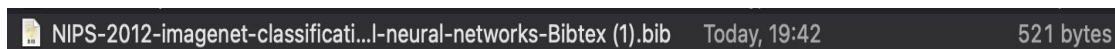
3. (Optional) Rename all files



4. Take notes + think about ideas



5. Cite them in your own paper



Check the Folder

Name	Date Modified	Size	Kind
AlexNet Paper, NIPs NIPS-2012-imagenet-classification-with-deep-convolutional-neural-networks-Paper.pdf	Today, 19:16	1,4 MB	PDF Document
N19-1423.pdf BERT paper, ACL	Today, 19:16	786 KB	PDF Document
3295222.3295349.pdf Attention is all you need, also NIPs, but file name is different	Today, 19:17	533 KB	PDF Document
2304.02643.pdf Segment Anything, arxiv version (with notes) Kirillov_Segment_Anything_ICCV_2023_paper.pdf and ICCV version (newer, without notes)	Today, 19:17 Today, 19:18	15,6 MB 11 MB	PDF Document PDF Document
1912.02047.pdf	Today, 19:23	2,8 MB	PDF Document
2301.11316.pdf	Today, 19:23	587 KB	PDF Document
2002.07526.pdf	Today, 19:23	2,3 MB	PDF Document
1308.0850.pdf A list of arxiv papers need to read	Today, 19:24	3,5 MB	PDF Document

Additional Problems:

Citation files, Paper tags, Paper storage, Cross-devices access, Notes summary, Related Papers, Updated version of a paper, ..., etc.

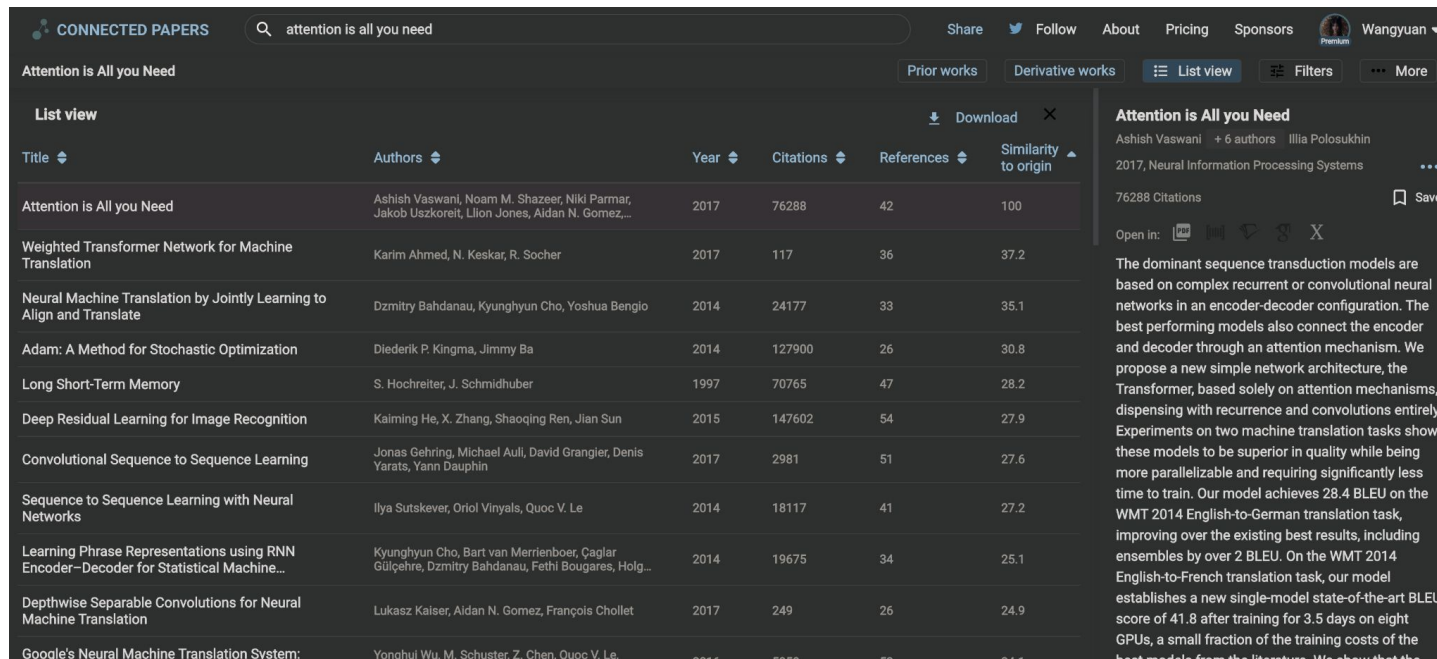


Want Better Solution?

Yes / No-Problem

Papers for Research (My Workflow)

1. Search Paper name on Connected Papers



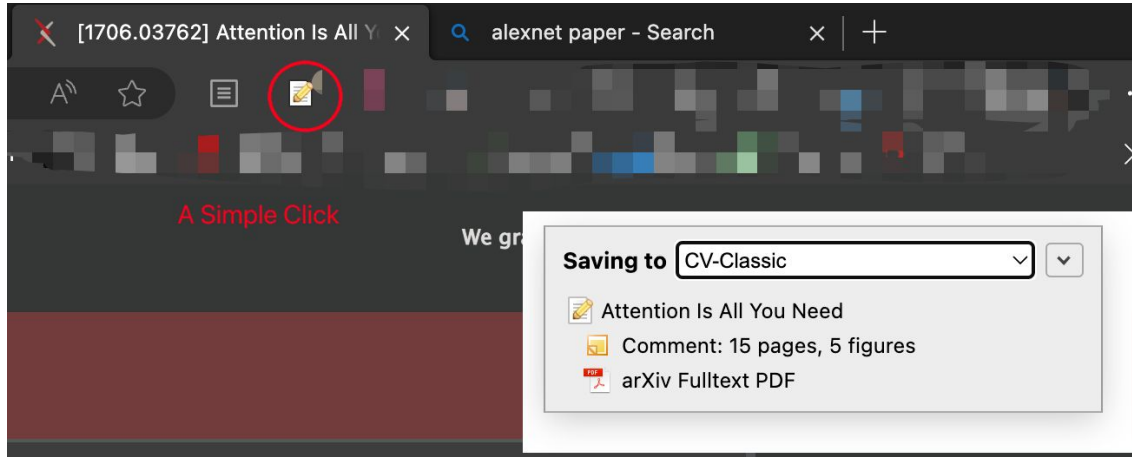
The screenshot shows the Connected Papers interface. At the top, there is a search bar with the query "attention is all you need". Below the search bar, there are navigation options: "Share", "Follow", "About", "Pricing", "Sponsors", and a user profile for "Wangyuan". The main content area displays a list of papers under the heading "List view". The papers are sorted by similarity to the origin, with the top paper being "Attention is All you Need" by Ashish Vaswani et al. (2017). A detailed view of this paper is shown on the right side of the screen, including its title, authors, year, and a brief description of the model.

Title	Authors	Year	Citations	References	Similarity to origin
Attention is All you Need	Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez,...	2017	76288	42	100
Weighted Transformer Network for Machine Translation	Karim Ahmed, N. Keskar, R. Socher	2017	117	36	37.2
Neural Machine Translation by Jointly Learning to Align and Translate	Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio	2014	24177	33	35.1
Adam: A Method for Stochastic Optimization	Diederik P. Kingma, Jimmy Ba	2014	127900	26	30.8
Long Short-Term Memory	S. Hochreiter, J. Schmidhuber	1997	70765	47	28.2
Deep Residual Learning for Image Recognition	Kaiming He, X. Zhang, Shaoqing Ren, Jian Sun	2015	147602	54	27.9
Convolutional Sequence to Sequence Learning	Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann Dauphin	2017	2981	51	27.6
Sequence to Sequence Learning with Neural Networks	Ilya Sutskever, Oriol Vinyals, Quoc V. Le	2014	18117	41	27.2
Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine...	Kyunghyun Cho, Bart van Merriënboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger...	2014	19675	34	25.1
Depthwise Separable Convolutions for Neural Machine Translation	Lukasz Kaiser, Aidan N. Gomez, François Chollet	2017	249	26	24.9
Google's Neural Machine Translation System:	Yonghui Wu, M. Schuster, Z. Chen, Quoc V. Le,...	2016	5950	53	24.3

Attention is All you Need
Ashish Vaswani + 6 authors Illia Polosukhin
2017, Neural Information Processing Systems
76288 Citations
Open in: PDF, HTML, etc.
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the

Papers for Research (My Workflow)

2. Select papers I am interested, automatically go to the paper website, Click add.



DONE!

Everything else can be done easily through Zotero!

Zotero

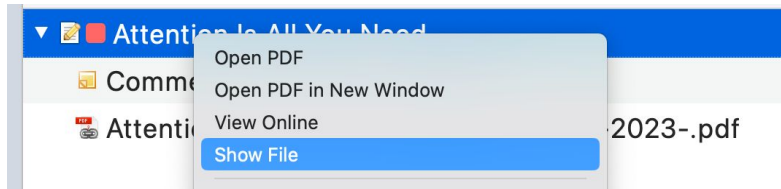
The screenshot displays the Zotero desktop application interface. On the left, a sidebar shows a hierarchical library structure under 'My Library', with a folder named 'CV-Classico' circled in red. A red annotation points to this folder: 'A Folder for papers of the same type'. The main pane shows a list of items with columns for Title, Year, Creator, Date Added, Rights, and Citation key. One item is selected, and a blue box highlights its title and file name. A blue annotation states: 'Automatically rename the pdf file'. Below the list, a search bar contains several tags, with a purple box around them and a purple annotation: 'Tags added to the paper, can search through them'. On the right, a detailed view of the selected item is shown, with a green box around it and a green annotation: 'All infos can be exported as a bibtex file Or we can export the whole folder'. The detailed view includes fields for Item Type (Preprint), Title (Attention Is All You Need), Author (Vaswani, Ashish), Date (2023-08-01), and various identifiers like DOI and arXiv ID.

Title	Year	Creator	Date Added	Rights	Citation key
Attention Is All You Need	2023	Vaswani et al.	12/12/2023, 20:00:50	★★★★★	vaswani2023Attention
Comment: 15 pages, 5 figures					
Attention_Is_All_You_Need@VaswaniA-2023-.pdf			12/12/2023, 20:00:50		
			12/12/2023, 20:01:00		

Item Type: Preprint
Title: Attention Is All You Need
- Author: Vaswani, Ashish
- Author: Shazeer, Noam
- Author: Parmar, Niki
- Author: Uszkoreit, Jakob
- Author: Jones, Llion
3 more...
(...) Abstract: The dominant sequence t...
Genre:
Repository: arXiv
Archive ID: arXiv:1706.03762
Place:
Date: 2023-08-01 y m d
Series:
Series Number:
DOI: 10.48550/arXiv.1706.037...
Citation Key: vaswani2023Attention
URL: http://arxiv.org/abs/1706...
Accessed: 12/12/2023, 20:00:50
Archive:
Loc. in Archive:
Short Title:
Language: en
Library Catalog: arXiv.org
Call Number:
Rights: ★★★★★
Extra: /Junread
Date Added: 12/12/2023, 20:00:50
Modified: 12/12/2023, 20:02:09

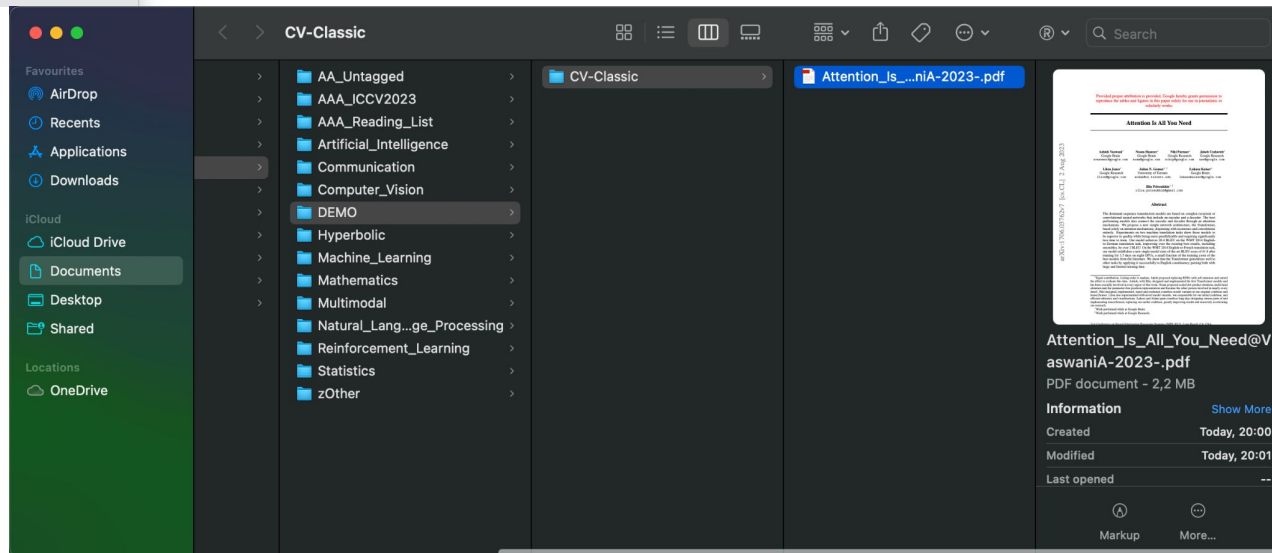
contrastive learning
cross-modal retrieval dataset
diffusion distillation gpt
graph neural network
image-text Machine learning
mask object discovery
self-supervised learning
set prediction
similarity reasoning

Where are My Files?



Show local PDF files

Same as in Zotero!
Can be on any
WebDev (iCloud or
OneDrive), and
okay to use any pdf
editors or note
apps!



Note Taking using PDF-Expert (or Adobe)

arXiv:2211.16761v1 [cs.LG]

..., which is designed to alleviate the side effects of existing similarity functions for set-based embedding. Second, we propose a novel set prediction module to produce a set of embedding vectors that effectively captures diverse semantics of input by the slot attention mechanism. Our method is evaluated on the COCO and Flickr30K datasets across different visual backbones, where it outperforms existing methods including ones that demand substantially larger computation at inference.

1. Introduction

Cross-modal retrieval is the task of searching for data relevant to a query from a database when the query and database have different modalities. While it has been studied for various pairs of modalities such as video-text [3, 7, 19] and audio-text [5, 16], the most representative setting for the task is the retrieval across image and text modalities [10, 30, 38, 47]. A naïve solution to cross-modal retrieval is a straightforward extension of the conventional unimodal retrieval framework [17, 18, 27], i.e., learning a joint embedding space of the different modalities with known ranking losses (e.g., contrastive loss [9] and triplet loss [3]). In this framework, each sample is represented as a single embedding vector and the task reduces to neighbor search on the joint embedding space.

However, this naïve approach has trouble in handling the inherent ambiguity of the cross-modal retrieval across image and text modalities [10, 30, 47]. A cause of the ambiguity is the fact that even a single image often contains various situations and contexts. Consider an image in Figure 1, which illustrates a group of children in a skate park. One of the captions coupled with it could be about children carrying up a bike, while another may describe the child riding a skateboard. Indeed, different local features of the image are matched to different captions. Similarly, visual manifestations of a caption could vary significantly as text descriptions are highly abstract. This ambiguity issue suggests that a sample should be embedded while reflecting its varying semantics in cross-modal retrieval. Embedding models dedicated to the uni-modal retrieval do not meet this requirement since they represent a sample as a single embedding vector.

Various methods have been studied to mitigate the ambiguity issue of cross-modal retrieval. Most of them adopt cross-attention networks that directly predict the similarity of an input image-caption pair [12, 14, 25, 31, 38, 39, 50, 51, 54]. These models successfully address the ambiguity since they explicitly infer relations across the modalities by drawing attentions on both modalities at once. However, they inevitably impose a large computation burden on retrieval systems since they demand both image and caption to be processed together for computing their similarity; all data in a database thus have to be reprocessed whenever a query arrives. In contrast, methods using separate textual and visual encoders [17, 21, 24, 30] seek to find sam-

Annotations:

- A A naïve solution to cross-modal retrieval is a straightforward extension of the conventional unimodal retrieval framework [17, 18, 27], i.e., learning a joint embedding space of the different modalities with known ranking losses (e.g., contrastive loss [9] and triplet loss [3]). In this framework, each sample is represented as a single embedding vector and the task reduces to neighbor search on the joint embedding space.
- A However, this naïve approach has trouble in handling the inherent ambiguity of the cross-modal retrieval across image and text modalities [10, 30, 47].
- A Indeed, different local features of the image are matched to different captions. Similarly, visual manifestations of a caption could vary significantly as text descriptions are highly abstract.
- A This ambiguity issue suggests that a sample should be embedded while reflecting its varying semantics in cross-modal retrieval.
- A Various methods have been studied to mitigate the ambiguity issue of cross-modal retrieval.
- A 两种方法:
1. Cross-attention networks, 优点: 直接计算similarity, 不存在ambiguity问题
缺点: 计算量太大
2. 分开的encoder
优点: 从计算好的vector embedding中找nearest neighbour来报导, 速度很快
缺点: 只有单一的vector embedding, 所以没法处理ambiguity问题
- A However, they inevitably impose a large computation burden on retrieval systems since they demand both

Automatically Gather Notes from PDF

The screenshot displays a library interface with a sidebar on the left containing a tree view of categories such as 'My Library', 'Computer_Vision', 'DEMO', 'Machine_Learning', 'Mathematics', 'Multimodal', and 'Natural Language Process'. The main area shows a table of research papers with columns for Title, Year, Creator, Date Added, Rights, and Citation key. The selected paper is 'Improving Cross-Modal Retrieval with Set of Diverse Embeddings' by Kim et al., dated 2023.

Title	Year	Creator	Date Added	Rights	Citation key
Improving Cross-Modal Retrieval with Set of Diverse Embeddings	2023	Kim et al.	23/10/2023, 16:51:52		kim2023Improving
注释笔记 W9GBW2QQ>>>>>>>>>>			16/12/2022, 12:02:36		
Comment: Accepted to CVPR 2023 (Highlight)			23/10/2023, 16:51:52		
Improving_Cross-Modal_Retrieval_with_Set_of_Diverse_Embeddings@KimD-2023-...			07/12/2022, 04:21:24		
Improving_Cross-Modal_Retrieval_with_Set_of_Diverse_Embeddings@KimD-2023-...			23/10/2023, 16:51:56		
Prototype-based Dataset Comparison	2023	van Noord	19/10/2023, 16:21:20		vannoord2023Protot...
VALOR: Vision-Audio-Language Omni-Perception Pretraining Model and Dataset	2023	Chen et al.	22/09/2023, 20:55:53		chen2023VALOR
DataComp: In search of the next generation of multimodal datasets	2023	Gadre et al.	19/09/2023, 01:08:03		gadre2023DataComp
LAION-400M: Open Dataset of CLIP-Filtered 400 Million Image-Text Pairs	2021	Schuhmann et...	19/09/2023, 01:06:44		schuhmann2021LAI...
Improving Multimodal Datasets with Image Captioning	2023	Nguyen et al.	17/09/2023, 21:15:40		nguyen2023Improving
Conceptual Captions: A Cleaned, Hypernymed, Image Alt-text Dataset For Automat...	2018	Sharma et al.	11/09/2023, 18:35:22		sharma2018Concept...
ChinaOpen: A Dataset for Open-world Multimodal Learning	2023	Chen et al.	15/08/2023, 18:47:57		chen2023ChinaOpen
GLIGEN: Open-Set Grounded Text-to-Image Generation	2023	Li et al.	15/06/2023, 15:19:23		li2023GLIGEN
LAION-5B: An open large-scale dataset for training next generation image-text mo...	2022	Schuhmann et...	24/04/2023, 14:39:54		schuhmann2022LAI...
Does progress on ImageNet transfer to real-world datasets?	2023	Fang et al.	16/01/2023, 01:35:59		fang2023Does
InPars-v2: Large Language Models as Efficient Dataset Generators for Information ...	2023	Jeronymo et al.	13/01/2023, 16:58:52		jeronymo2023InPars...
Telling the What while Pointing to the Where: Multimodal Queries for Image Retrieval	2021	Changpinyo e...	18/12/2022, 22:47:49		changpinyo2021Tell...

The right-hand pane shows a detailed view of the selected paper, including the title '注释笔记 W9GBW2QQ>>>>>>>>>>', a citation key 'kim2023Improving', and a list of notes extracted from the PDF. The notes discuss a naive solution to cross-modal retrieval, the inherent ambiguity in handling local features of images, and various methods studied to mitigate this ambiguity issue.

Notes extracted from the PDF:

- retrieval is a straightforward extension of the conventional unimodal retrieval framework [17, 18, 27]. I.e., learning joint embedding space of the different modalities with known ranking losses (e.g., contrastive loss [9] and triplet loss [45]). In this framework, each sample is represented as a single embedding vector and the task reduces to neighbor search on the joint embedding space. (n)
- trouble in handling the inherent ambiguity (n)
- Indeed, different local features of the image are matched to different captions. Similarly, visual manifestations of a caption could vary significantly as text descriptions are highly abstract. (n)
- This ambiguity issue suggests that a sample should be embedded while reflecting its varying semantics in cross-modal retrieval. (n)
- Various methods have been studied to mitigate the ambiguity issue of cross-modal retrieval.
- 两种方法:
 - Cross-attention network. 优点: 直接计算 similarity, 不存在 ambiguity 问题. 缺点: 计算量大
 - 分开的 encoder. 优点: 从计算好的 vector embedding 中按 nearest neighbour 来搜索, 速度很快. 缺点: 只有单一的 vector embedding, 所以没法处理 ambiguity 问题. (n)
- all data in a database thus have to be reprocessed whenever a query arrives. (n)
- nearest-neighbor search on pre-computed embedding vectors (n)

Related: [click here]
Tags: /注释笔记

Advanced Play (Extract all Infos and take notes)

A Improving Cross-Modal Retrieval with Set of Diverse Embeddings

Metadata

Key	Value
Type	preprint
Title	Improving Cross-Modal Retrieval with Set of Diverse Embeddings
Short	
Authors	Dongwon Kim, Namyup Kim, Suha Kwak
Journal Name	
DOI	
Place	
ArXiv	arXiv.org
???	???
CopyRights	
Class	AAA_Reading_List, Slot_Attention
Link	My Library
PDF Link	<ul style="list-style-type: none">Improving_Cross-Modal_Retrieval_with_Set_of_Diverse_Embeddings@KimD-2023-.pdfImproving_Cross-Modal_Retrieval_with_Set_of_Diverse_Embeddings@KimD-2023-.pdf
Related Files	2021_Ohun_Probabilistic_Embedd_KEY-CMADW4A2, 2023_Pel_Complementarity_Is_1_KEY-D8P2U2H, 2019_Song_Polysemous_Visual-Se_KEY-TXKYKN15

≡ 本文标签 >

🔍 Abstract

Cross-modal retrieval across image and text modalities is a challenging task due to its inherent ambiguity: An image often exhibits various situations, and a caption can be coupled with diverse images. Set-based embedding has been studied as a solution to this problem. It seeks to encode a sample into a set of different embedding vectors that capture different semantics of the sample. In this paper, we present a novel set-based embedding method, which is distinct from previous work in two aspects. First, we present a new similarity function called smooth-Chamfer similarity, which is designed to alleviate the side effects of existing similarity functions for set-based embedding. Second, we propose a novel set prediction module to produce a set of embedding vectors that effectively captures diverse semantics of input by the slot attention mechanism. Our method is evaluated on the COCO and Flickr30K datasets across various retrieval tasks. We compare our method with existing methods including ones that derive the similarity matrix by using the joint embedding space.

A.1 Notes

Summary

- Research Question: Set-based Model
- Background: Multi-view Learning and Object-centric Learning
- Methodology: Slot-attention
- Idea: Use the ability of the slot-attention module to deal with multi-view cross-modal retrieval
- Main Body: Same
- Conclusion: Achieve SOTA performance on MSCOCO and Flickr 30 K
- Innovation: First time apply slot-attention on the task of cross-modal retrieval
- Limitation: Still cant hold diverse captions
- Future Work: Improve over the slot-attention module, maybe probabilistic embedding?
- PS:

Import Date

importDate: 2023-11-26

ReadList

Title (1)	Tag	Rate	Zotero	Notes
Improving Cross-Modal Retrieval with Set of Diverse Embeddings	<ul style="list-style-type: none">#unread#Done#cross-modal_retrieval#slot_attention	-	My Library	2023_Kim_Improving_Cross-Modal_KEY-FQ9WBH8F

A.2 Notes W9GBW2QQ

↑ 2023-Kim-Improving Cross-Modal Retrieval with Set of Diverse Embeddings

^KEY9fTitle

A naive solution to cross-modal retrieval is a straightforward extension of the conventional unimodal retrieval framework [17, 18, 27], i.e., learning a joint embedding space of the different modalities with known ranking losses (e.g., contrastive loss [9] and triplet loss [45]). In this framework, each sample is represented as a single embedding vector and the task reduces to neighbor search on the joint embedding space. (p1)

^KEYLDKDTL9A

Success in handling the inherent ambiguity. (p1)

^KEYL3JKVBJS

Indeed, different local features of the image are matched to different captions. Similarly, visual manifestations of a caption could vary significantly as text descriptions are highly abstract. (p1)

^KEY2QYXH75

The ambiguity issue suggests that a sample should be embedded while respecting its varying. (p1)

^KEY6HQNK27X

Various methods have been studied to mitigate the ambiguity issue of cross-modal retrieval.

两种方法:

- Cross-attention network.
优点: 直接计算similarity, 不存在ambiguity问题
缺点: 计算量太大
- 分开的encoder
优点: 从计算好的vector embedding中按nearest neighbour来搜索, 速度很快
缺点: 只有单

!)

^KEYR54Z4A4D

Reference Management (Summary)

1. I use Zotero, but other softwares are also good
2. Simple click then papers (also forum posts, online article, etc.) go to zotero, with a pre-defined name format
3. Files can be stored using any WebDev
4. Extract notes, cross-device access, track papers read
5. Easily export citation of the whole project folder as bibtex
6. Archive the project folder once the project is complete, along with all notes

I will share the whole configuration (~5 minutes video) of zotero.

And after that, research becomes 5x easier and well-organized.

Random opinions

Cleanup your latex

google-research / arxiv-latex-cleaner

Code Issues 21 Pull requests 1 Actions Security Insights

arxiv-latex-cleaner Public Watch 29 Fork 288 Star 4.4k

main 1 Branch 30 Tags

Go to file Add file Code

tnaumann Adds Homebrew installation instructions (#64) 9f47c5a · last month 148 Commits

.github/workflows	PR cleanup	3 years ago
arxiv_latex_cleaner	Accept zip-compressed folders as input_dir (#83)	3 months ago
tex	Remove user-defined environments (#80)	6 months ago
tex_arXiv_true	Remove user-defined environments (#80)	6 months ago
.gitignore	PyPI GitHub Release deployment prep	3 years ago
CONTRIBUTING.md	arXiv LaTeX cleaner moved to its own repo from https://git...	5 years ago
LICENSE	arXiv LaTeX cleaner moved to its own repo from https://git...	5 years ago
MANIFEST.in	PyPI GitHub Release deployment prep	3 years ago
README.md	Minor change in the README	5 months ago
__init__.py	arXiv LaTeX cleaner moved to its own repo from https://git...	5 years ago
cleaner_config.yaml	v0.1.13	2 years ago

About

arXiv LaTeX Cleaner: Easily clean the LaTeX code of your paper to submit to arXiv

latex arxiv

Readme Apache-2.0 license Activity 4.4k stars 29 watching 288 forks Report repository

Releases 30

Release v1.0.1 Latest on Jul 4

+ 29 releases

Clean up bib file to only have things you referenced



With a TeX Live distribution (possibly also with MiKTeX) there is a `bibexport` program. Assuming your document is `myarticle.tex`, you have to compile it normally and then you call

150

```
bibexport -o extracted.bib myarticle.aux
```



where `extracted.bib` is the name that you want to give to your new `.bib` file. Notice that you have to give the extension `.aux` (or no extension at all).



Then you have to change the name of the `.bib` file in your document, in order to use `extracted.bib`.

Share Improve this answer Follow

edited Jan 21, 2012 at 15:33

answered Jan 21, 2012 at 15:05



egreg

1.1m ● 135 ● 2615 ●

4215