

Toward Related Work Generation with Structure and Novelty Statement

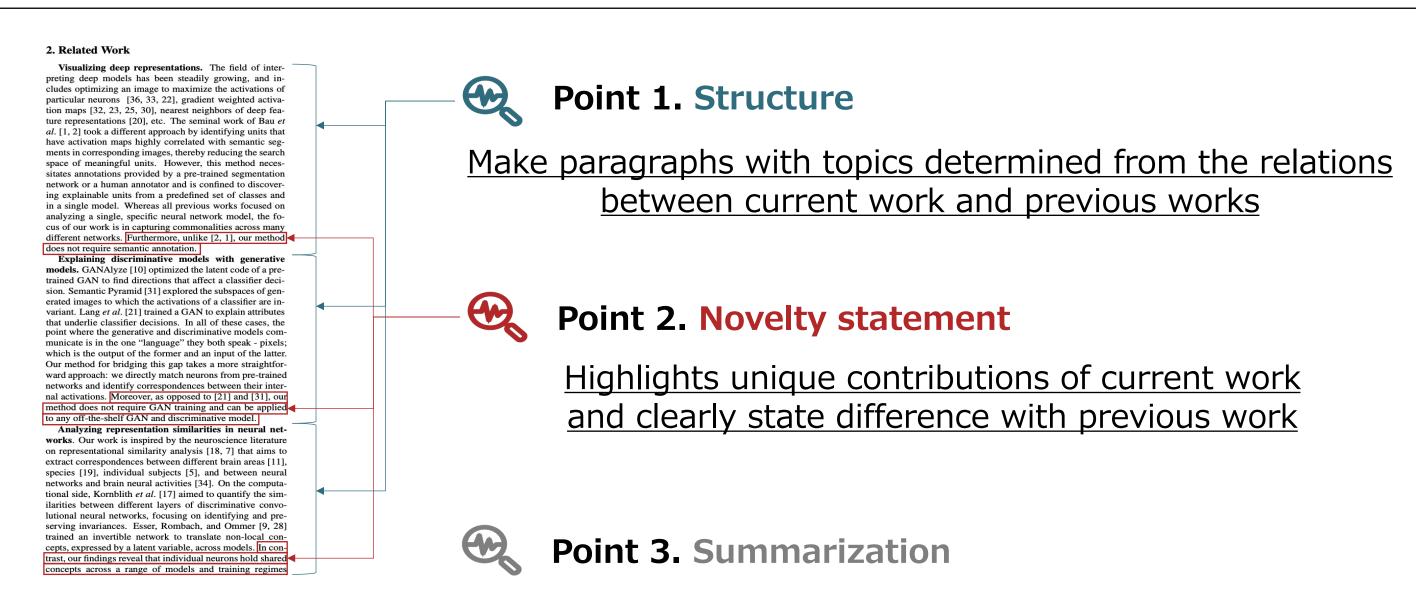
Scholarly Document

Processing @ACL2024

SINICIC

Kazuya Nishimura¹, Kuniaki Saito², Tosho Hirasawa², Yoshitaka Ushiku² National Cancer Center Japan, OMRON SINIC X Corp.

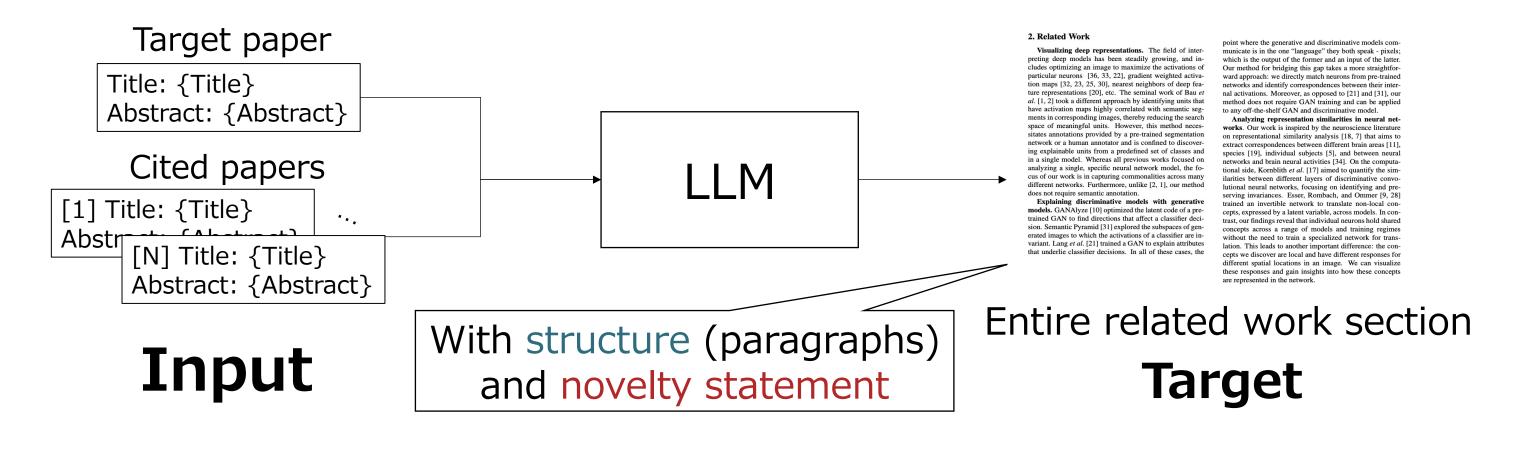
What is well-written related work?



While structure and novelty statement are important for related work, previous work has treated automated related work generation as a just summarization

STructured Related work Generation with Novelty Statemen (STRoGeNS)

Redefine related work generation task as **STRoGeNS**



Problems. Lack of datasets and metrics

Problem 1. no datasets for STRoGeNS

Dataset	Target	Pairs	#Para.
Multi-Xscience [Lu+, 2020]		40,528	1
S20RC [Chen+, 2021]		136,655	1
Delve [Chen+, 2021]	Only 1 paragraph in related work section	78,927	1
TAS2 [Chen+, 2022]		117,700	1
TAD [Chen+, 2022]		218,255	1
BigSurvey-MDS [Liu+, 2022]		4,478	1
SciReviewGen [Kasanishi+, 2022]		10,130	1

Problem 2. no evaluation metrics for STRoGeNS

Solutions. Propose datasets and metrics

Solution 1. Propose large-scale datasets for that took into consideration both quality and quantity.

	_		<u> </u>
Dataset	Target	Pairs	#Para.
STRoGeNS-arXiv22 (quantity)	Cturetryed veleted week	85,853	4.22
STRoGeNS-conf22 (quality)	Structured related work	15,079	4.27
STRoGeNS-conf23 (Test)	with novelty statement	4,762	4.04

Solution 2. Propose automated evaluation metrics for structure and novelty.

Quality vs quantity

Method	Dataset	R1 ↑	F1 ↑
BART	conf22	42.5	62.5
[Lewis+, 2019]	arXiv22	46.1	66.4
PEGASUS	conf22	18.1	31.2
[Zhang+, 2020]	arXiv22	17.7	40.2
LED	conf22	41.5	56.1
[Beltagy+, 2020]	arXiv22	40.9	12.3
Llama2-7B [Touvron+, 2023]	conf22	42.6	70.0
	arXiv22	37.6	9.0

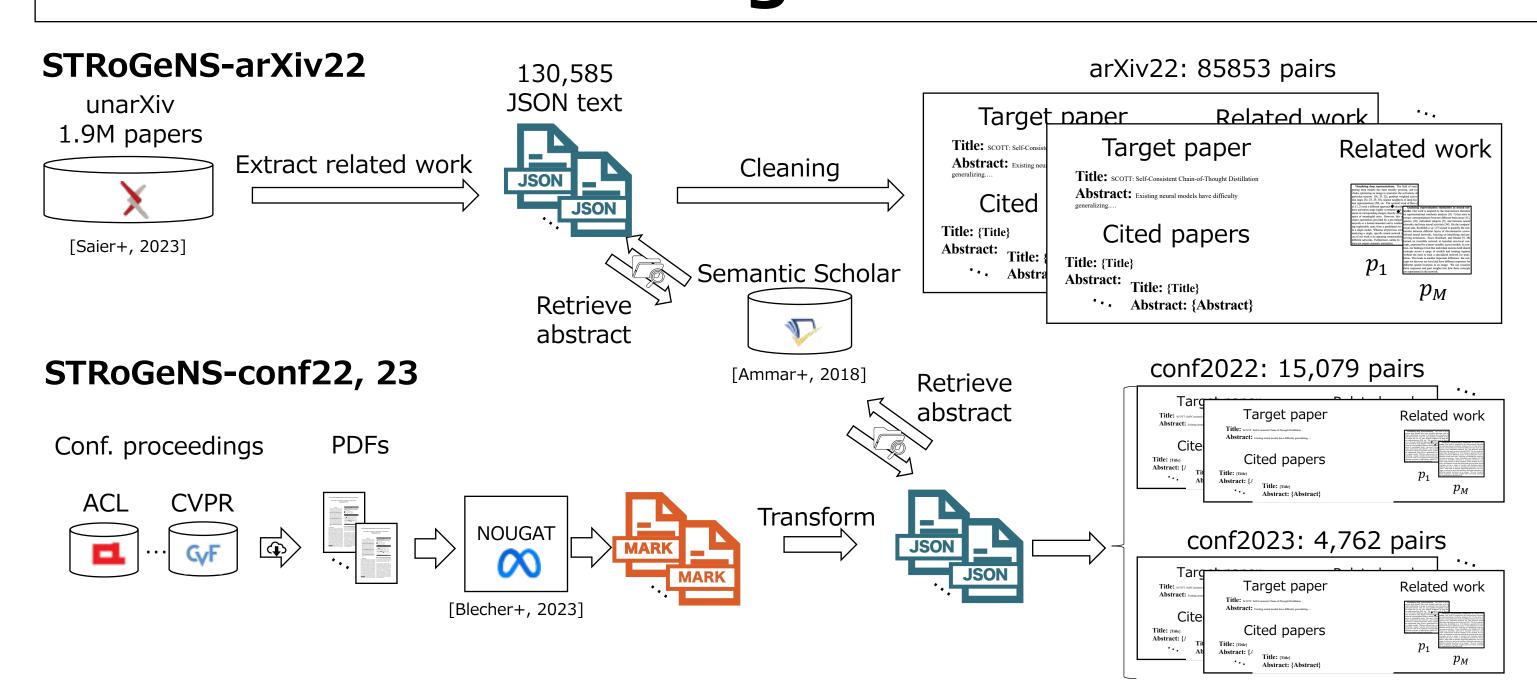
Novelty statement evaluation

Method	Novelty score
GPT3.5	0.17
BART	0.91
LED	0.53
Llamma2	0.56

Human evaluation

Metrics	r	$oldsymbol{ ho}$	au
F1	47.2	56.4	31.9
ARI′	38.3	35.7	24.9

Dataset generation

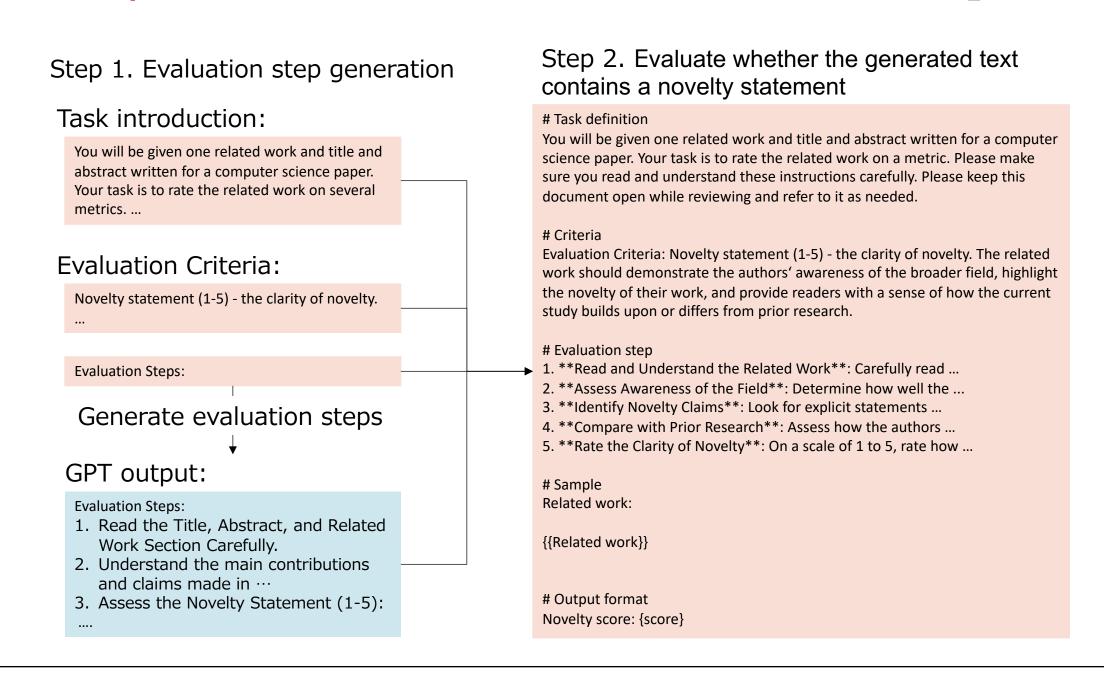


Evaluation metrics

Structure evaluation based on citation trends

Motiv	ation				Metrics
Evaluate how similar a of generated results		Ref. ID		ter ID Gen.	F1 measure citation similarity among most similar $F1 = \frac{1}{m} \sum_{i=1}^{m} \max_{j} F1(f(p_i), f(j))$
Gold standard	Generated	[1]	1	1	i=1
Gold Stalldard	Generated	[2]	1	1	ADT avaluate situation transfer like eluctoring avaluation
$p_1 = [1], [2], [3]$	$\hat{p}_1 - \frac{[1],[2]}{[1]}$	[3]	1	UC	ARI evaluate citation trends like clustering evaluation
$p_1 \underline{-[1],[2],[3]}$	$p_1 = \frac{1}{1}$	[1]	UC	2	#TPair
		[3]	2	2	$ARI' = \frac{\#TPair}{\#Pair - \#UC}$
$p_2 = -[3], [4], [5]$	\hat{p}_2 [1],[3],[4]	[4]	2	2	DI/ Even act ad DI
		[5]	2	MC	$ARI' = \frac{RI' - ExpectedRI}{max(RI) - ExpectedRI}$

Novelty statement evaluation with G-Eval [Liu et al. 2023]



Accuracy of novelty score is 90%.

Experiments

Structure and summarization comparisons

			•		
Mothod	Stru	cture	Su	ion	
Method	F1 ↑	ARI′↑	R1↑	R2 ↑	R-L↑
TextRank [Mihalcea+, 2004]	-	-	31.6	7.5	24.7
LexRank [Erkan+, 2021]	-	_	31.5	8.19	21.6
GPT3.5 [Brown+, 2020]	69.0	33.9	38.2	7.7	28.5
GPT4 [Brown+, 2020]	66.7	29.8	41.4	7.8	29.8
BART [Lewis+, 2019]	69.1	54.0	47.4	13.1	32.7
PEGASUS [Zhang+, 2020]	48.8	11.0	25.4	6.2	19.1
LED [Beltagy+, 2020]	59.5	33.7	44.0	11.8	29.2
Llama2-7B [Touvron+, 2023]	69.8	47.3	42.7	9.0	29.4

Example of generated related work section

Title: Space Engage: Collaborative Space Supervision for Contrastive-based Semi-Supervised Semantic Segmentation **Abstract:** Semi-Supervised Semantic Segmentation (S4) aims to train a segmentation model with limited labeled images and a substantial volume of unlabeled images. To improve the GANs [5], adversarial training [6], and consistency robustness of representations, powerful methods introduce a pixel-wise contrastive learning approach in latent space (i.e., representation space) that aggregates the representations to their prototypes in a fully supervised manner. However, previous contrastive-based S4 methods merely rely on the supervision from the model's output (logits) in logit space during unlabeled training. In contrast, we utilize the outputs in both logit space and representation space to obtain supervision in a collaborative way. The supervision from two spaces plays two roles: 1) reduces the risk of over-fitting to incorrect semantic information in logits with the help of representations; 2) enhances the knowledge exchange between the two spaces. training and additionally explore semantic information among Furthermore, unlike previous approaches, we use the similarity different images. between representations and prototypes as a new indicator to tilt training those under-performing representations and achieve Pixel-wise contrastive learning explores semantic relations not only in the individual image but also among different images. a more efficient contrastive learning process. Results on two Different from instance-wise contrastive learning [24][25][26], pixel-wise contrastive learning [27][28][29][30] project each pixel to the representation in representation space with the

public benchmarks demonstrate the competitive performance of our method compared with state-of-the-art methods. [1] Adversarial Learning for Semi-supervised Semantic Segmentation, Abstract: We propose a method for semisupervised semantic segmentation using an adversarial network. While most existing discriminators are trained to classify input images as real or fake on the image level, we design a discriminator in a fully convolutional manner to differentiate the predicted probability maps from the ground truth segmentation distribution with the consideration of the spatial resolution. We show that the proposed discriminator can be used to improve semantic segmentation accuracy by coupling the adversarial loss with the standard cross entropy loss of the proposed model. In addition, the fully convolutional discriminator enables semi-supervised learning through contrastive learning progress. discovering the trustworthy regions in predicted results of unlabeled images, thereby providing additional supervisory signals. In contrast to existing methods that utilize weakly-

learning [35][36][37] and unsupervised domain adaption labeled images, our method leverages unlabeled images to enhance the segmentation model. Experimental results on the PASCAL VOC 2012 and Cityscapes datasets demonstrate the classifier [43]. Concretely, the classes in the dataset are effectiveness of the proposed algorithm. presented by a set of non-learnable prototypes, and the dense semantic predictions are thus achieved by assigning the output features to its most similar prototype. Under semi-supervised '[25] A Simple Framework for Contrastive Learning of Visual Representations, Abstract: This paper presents SimCLR: a settings, some methods maintain the consistency between simple framework for contrastive learning of visual predictions from a linear predictor and a prototype-based representations. We simplify recently proposed contrastive selfpredictor. The two predictors are followed by the encoder and supervised learning algorithms without requiring specialized project the features to logit space and representation space,

architectures or a ... accuracy, which is a 7% relative

improvement over previous state-of-the-art, matching the

performance of a supervised ResNet-50. When fine-tuned on

The aim of S4 is to train a segmentation model with the semiSemi-Supervised Semantic Segmentation. Semi-supervised Semi-supervised Semantic Segmentation (SSL) aims to supervised setting (i.e., a few labeled images and a large number of unlabeled images) to classify each pixel in an entire unlabeled data to improve the segmentation performance of substantial volume of unlabeled data. Adversarial-based image. The critical issue of S4 is how to leverage unlabeled images to train the model. Some methods [1][2][3][4] based on regularization paradigm [7][8][9][10]. Meanwhile, self-training [11][12][13][14][15] is also a striking paradigm, which always generates pseudo-labels from model and retrains the model with the combined supervision of human annotations and pseudolabels. One essential issue of self-training is the accuracy of pseudo-labels. Some methods [16][17][18][19] try to polish pseudo-labels and provide reliable guidance. Some methods [20][21][22][23] focus on the class-imbalance problems in the dataset and try to alleviate the negative effect from class-biased pseudo-labels generated by the model pre-trained on imbalanced with other traditional S4 methods labeled images. We build our framework based on the self-

[13][12][19][20][21][22][23]. However, previous self-training methods mainly rely on the learning has been widely used in unsupervised visual representation learning to disentangle visual features by utilizing strong supervision [24][25][26][27]. Recently, cooperation of encoder and representation head. Representations learning to improve the representations of images and their results. PiPa [29] and PiPaV2 [27] perform pixel-wise and are then aggregated in their prototypes and are separated from associated pseudo labels [28][29][30][31][32]. each other in different classes. In semi-supervised settings, most methods [31][32][33][34] use pseudo-labels based on logits to provide semantic information contrastive learning process during labels to form an augmented input dataset to supplement the supervise the student model, which is similar to Meantraining on unlabeled images. Meanwhile, the confidence of logit limited labeled data and train the segmentation model with Teacher [12].

is used as an indicator to involve the contrastive learning process, e.g., [32] uses the hard representations whose corresponding logit confidence is lower than a threshold to contrast for effective training. As opposed to the above methods, we use collaborative space supervision for contrastive learning on unlabeled images and use a new indicator to involve the Prototype-based learning has been widely studied in few-shot [38][39][40][41][42]. Recently, it is restudied in semantic segmentation as known as a non-parametric prototype-based

respectively. In this work, we combine the semantic information

in the logit and representation spaces to provide supervision in a

collaborative way during semi-supervised learning.

this augmented dataset. An Augmented Learning-based Pixel-

feature representation, but it is designed to work with a fully- domain adaptation. PCL [34] is the first work to use supervised image labeling scheme in the segmentation inputs used in ALP are obtained with the ground truth labels. contrast, we utilize the outputs in both logit and S4[33] simply selects highly confident pixels as pseudo labels representation spaces to supervize each other. and updates the model with the augmented dataset, thus does not have similar supervision as ALP. In contrast, C3-SS [14] proposes a contrast-based method to learn discriminative trains a segmentation model with pseudo labels in logit space features for image classification. SimCLR [25] proposes a and the ground-truth labels in representation space. The features extracted from the logit-space segmentation model latent space. Inspired by these works, we propose a pixel-

and prototypes with an extra representation-to-prototype

the network outputs in

of S4 models is first to predict the probabilities for the

model with these pseudo labels and real labeled images

[7][8][9][10][11]. Although robust models trained with S4

real-world applications due to the insufficient supervision

in segmentation accuracy and temporal stability compared supervised learning and consistency regularization. ST++ [19] improves FixMatch by applying multi-level perturbations. CReST [21] uses class-rebalancing to improve the quality of fully-supervised model outputs in logits space to improve the Contrastive-based Learning [24][25][26] has achieved great robustness of feature representations. In contrast, contrastive success in unsupervised visual representation learning. Recently, contrastive learning has also been introduced to SSL. C3-SemiSeg [14] uses a contrastive loss in the pixellevel. Pixel-wise contrastive methods

labels for unlabelled data. VAT [6] and CCT [7] use

consistent with the ground truth. UCC [8] and CPS [9] use

uncertainty estimation to enhance pseudo-label generation

framework that combines a teacher model and a student

contrastive learning has been introduced to semi-supervised [27][28][29][30][31][32][33][34] also show promising feature-level contrastive losses. PCC [30] and PCC-v2 [31] use a consistency loss between global and local features. UCT Most of these contrastive-based methods introduce the pseudo [33] uses unreliable predictions from the teacher model to

semantic segmentation (S4) aims to leverage abundant learn a segmentation model with limited labeled data and a

limited labeled images [1][2][3][4][5][6]. The basic workflow methods [1][2][3][4] utilize GANs [5] to generate pseudo-

unlabeled images (assigning pseudo labels) and then train the adversarial training to encourage the predictions to be

have achieved outstanding performance, it still struggles in Transformer-CNN Cohort [10] introduces a self-training

years, self-training (ST) based methods have emerged as a [11][12][13][14][15][16][17][18][19][20][21][22][23] are

popular approach and has achieved significant improvements also widely used in S4. FixMatch [18] combines self-

from the labeled data [12][13][14][15][16][17][18]. In recent model. Self-training methods

wise Contrastive Framework (ALP) introduces a contrastive Prototype-based learning [35][36][37][38][39][40][41][42] loss and a representation loss to improve the robustness of has been applied to many tasks, such as few-shot learning and prototypes to boost the performance of SSL. However, PCL pipeline (i.e., images are labelled before training). Also, the only uses the supervision from the output in logit space. In

simple framework that performs instance discrimination in the are fed into a contrastive loss to enhance their robustness. We wised contrastive method for S4, which differs from previous also use the features in logit space to produce a contrastive works in two aspects. First, we use the similarity between loss, but unlike previous contrastive-based S4 methods, we representations and prototypes as a new indicator to tilt propose to enhance the similarity between the representations training those under-performing representations. Second, we enhance the knowledge transfer between the two spaces by contrastive loss. Furthermore, unlike C3-SS which only uses introducing a collaborative learning strategy.

> introduces a contrast loss to learn transferable features in a fully supervised manner. [43] extends this idea to S4 and proposes a dual-path self-distillation method to improve