(a) Framework Architecture

L^{CLS} L^{SC}, L^{KP}

negative pair $\tilde{z}_{j \neq i}$

positive pair

Classifier

Representation z_i $\tilde{z}_i$

Adapter

Hidden States x_i $\tilde{x}_i$

PLM

Tokens

Input Template

Original Sample T_i $\xrightarrow{\text{SC or KP}}$ Transformed Sample

(b) Sentence-Level Cropping (SC)

Sequential Order T_i

...

-1 | As computer vision techniques have improved over the past decade, other research has begun directly using visual information in place of feature norms.

0 | The first work to do this with topic models is [#CITATION_TAG].

1 | They use a Bag of Visual Words model to create a bimodal vocabulary describing documents.

2 | The topic model using the bimodal vocabulary outperforms a purely textual based model in word association and word similarity prediction.

3 | Bruni et al. (2012a) show how a BoVW model may be easily combined with a distributional vector space model of language.

...

$\tilde{T}_{i,0}^{SC}$
-1	-b
0	
1	v

$\tilde{T}_{i,1}^{SC}$
| 0 | -b |
| | v |

$\tilde{T}_{i,2}^{SC}$
0	-b
1	
2	
3	v

(c) Keyphrase Perturbation (KP)

*The first work to do this with <u>topic models</u> is [**#CITATION_TAG**].*

Lr   Technique

They use a <u>Bag of Visual Words model</u> to <mark>create</mark> a <u>bimodal vocabulary</u>

Technique   Concept

Synonym 🔍   **Gr** 🔍

describing <u>documents</u>.

Material   **Ab**

Word net   STK Base

*The first work to do this with <u>Bag of Visual Words model</u> is [**#CITATION_TAG**]. They use a <u>Bag of Visual Words model</u> to <mark>build</mark> a <u>semantic representation</u> describing <u>Material-1</u>.*