

Annotation Interface :

Header: True PwC Entity names that needs to be marked. In top 4 lines we show the true entity names that PwC provide us, categorised by Entity types (Material, Metric, Task, Method)

Method names can sometimes be composed of multiple parts where individual parts may be present as separate spans in document. We have used a heuristic method to split the model names into their subparts (If the splitting is wrong please delete current split annotations and resplit them).

If splitting or resplitting needs to be done on a model name, please annotate the correct split and move on to the next document. We will do a second pass of annotations for these documents.

1	MATERIAL: PROMISE 2012
2	METRIC: Dice Score
3	TASK: Volumetric Medical Image Segmentation
4	METHOD: V-Net + Dice-based loss

Content : This is the document to be annotated. The document contains spans of text marked with entity types. Some of the spans are also marked with red # symbol which denotes that the span matches one of the pwc entities (or one of the splits of pwc entities) - we call these spans as **linked** .

6	document : V - Net : Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation
7	Convolutional Neural Networks (CNNs) have been recently employed to solve problems from both the computer vision and medical image analysis fields .
8	Despite their popularity , most approaches are only able to process 2D images while most medical data used in clinical practice consists of 3D volumes .
9	In this work we propose an approach to 3D image segmentation based on a volumetric , fully convolutional , neural network .
10	Our CNN is trained end - to - end on MRI volumes depicting prostate , and learns to predict segmentation for the whole volume at once .
11	We introduce a novel objective function , that we optimise during training , based on Dice coefficient .
12	In this way we can deal with situations where there is a strong imbalance between the number of foreground and background voxels .
13	To cope with the limited number of annotated volumes available for training , we augment the data applying random non - linear transformations and histogram matching .
14	We show in our experimental evaluation that our approach achieves good performances on challenging test data while requiring only a fraction of the processing time needed by other previous methods .
15	section : Introduction and Related Work
16	Recent research in computer vision and pattern recognition has highlighted the capabilities of Convolutional Neural Networks (CNNs) to solve challenging tasks such as classification , segmentation and object detection , achieving state - of - the - art performances .
17	This success has been attributed to the ability of CNNs to learn a hierarchical representation of raw input data , without relying on handcrafted features .

Edit Annotation

Text
V - Net
Link

Search
Google, Wikipedia

Entity type
☐ Material
☐ Metric
☐ Task
☒ Method

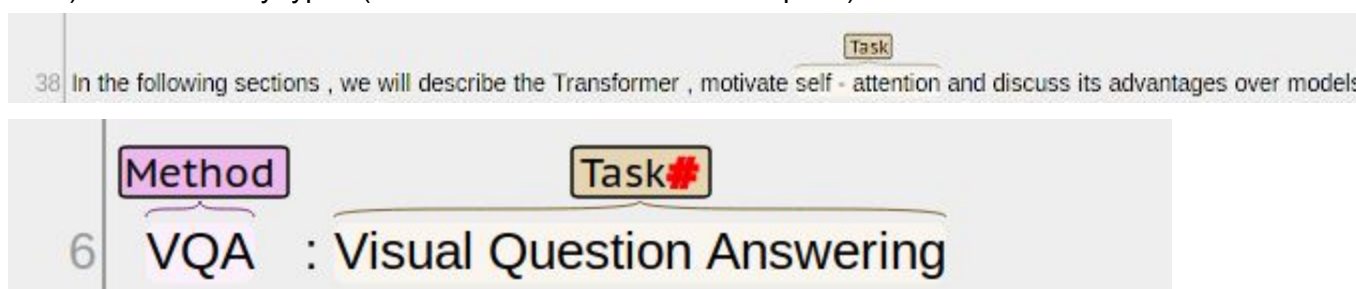
Entity attributes
☒ Canonical Name
☐ V-Net + Dice-based loss
☒ V-Net
☐ Dice-based loss

Notes

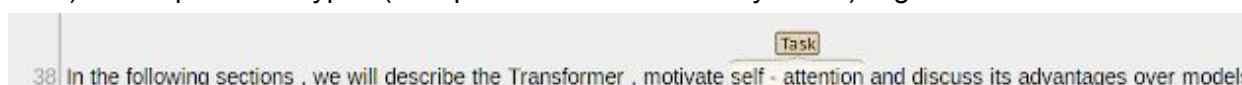
Add Frag.
Delete
Move
OK
Cancel

Entity Typing :

- 1) Correct entity types (for both matched / unmatched spans).

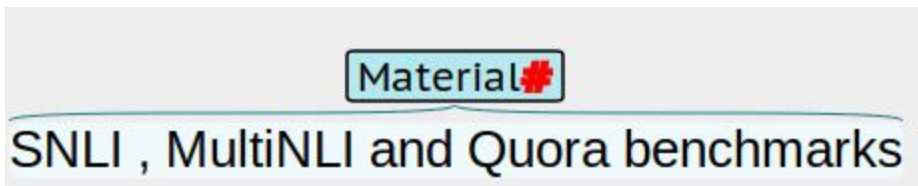


- 2) Delete Spans that are incorrectly marked as entity (both matched / unmatched spans).
- 3) Add spans and types (if a span match a PwC Entity name). Eg. Transformer below -



- 4) Important to focus on span that can be matched to one of PwC Entity names.

- 5) If a span contains multiple entities, split them into independent spans.



Type Specific Guidelines :

- 1) Datasets/Material span should be labeled as “Material” if they are being used and evaluated in the paper and/or refer to a named dataset.

PwC Entity Name Matching :

- 1) Each marked span can be matched to atmost one PwC entity name. We call these spans **linked** (marked with red #).. Please select the best match if multiple feasible are available.
- 2) **We know that some entities do not occur at all in the document so please make best effort to search for a span for each PwC entity name but it is NOT a requirement.**
- 3) Note: If a single entity can match multiple PwC Entity names, that implies that name may need to split into parts.
- 4) Canonical name: An extra attribute is present in annotation box to mark if the span correspond to PwC Entity name as canonical (since PwC Entity name may not occur exactly as provided, we want to find the best mention in the document of that PwC Entity) .
- 5) You can mark multiple spans in the document as canonical - we will consider the last span marked as canonical as true one.
- 6) Delete matched links if wrong for spans already annotated.
Example : Wrong Material Match

MATERIAL: Material COCO Visual Question Answering (VQA) Material real images 2.0 open ended |
Material COCO Visual Question Answering (VQA) Material abstract images 1.0 open ended |

Material Given an image and a natural language question about the image , the task

- 7) Add links to spans that match a PwC Entity name but not linked (important to take care of Abbreviations).

Method
a bi - directional attention flow mechanism :

8) Add Spans that are not already annotated if they match a PwC Entity name.

38 In the following sections , we will describe the Transformer , motivate self - attention and discuss its advantages over models

9) In case where same PwC Entity name appears in multiple types , please be careful to select the correct type for a span that matched that name. For example, below “Dice score” is both a method and a metric. The preexisting annotation marked “Dice Coefficient” as metric, but according to its context it is actually behaving as method. Therefore its type should be changes to Method and it should be linked to “Dice-based loss”

1 MATERIAL: PROMISE 2012 |

2 METRIC: Dice Score |

3 TASK: Volumetric Medical Image Segmentation |

4 METHOD: V-Net + Dice-based loss |

11 We introduce a novel objective function , that we optimise during training , based on Dice coefficient .

12 In this way we can deal with situations where there is a strong imbalance between the number of foreground

Metrics to keep track of:

1) Please keep track of time for each document annotation to be completed. We can keep using the previous spreadsheet and put in start and end timestamps for each document.