Ontology-driven Event Type Classification in Images
Supplemental Material

Eric Müller-Budack[1], Matthias Springstein[1], Sherzod Hakimov[1], Kevin Mrutzek[2], Ralph Ewerth[1,2]
[1]TIB - Leibniz Information Centre for Science and Technology, Hannover, Germany
[2]Leibniz University Hannover, L3S Research Center, Hannover, Germany
{eric.mueller, matthias.springstein, sherzod.hakimov, ralph.ewerth}@tib.eu

A. Overview

The supplemental material contains detailed information on the sampling strategy to download images (Section B) and dataset statistics for the train and test dataset (Section C). Furthermore, Web links to explore versions of the *Visual Event Ontology (VisE-O)* as well as the benchmark ontologies are provided (Section D) and results using different inference strategies (Section E) are presented.

B. Dataset Event Image Sampling

We described the data collection for the *Visual Event Classification Dataset (VisE-D)* in Section 3.3 of the paper. In this context, we have used a sampling strategy to download images for concrete events such as *Skeleton at the 2018 Winter Olympics* of a *Leaf Event Node* (in this case *Skeleton*) to increase the number of images and reduce ambiguities. The number of images downloaded for an *Event* is based on its popularity and date to prevent spam in the search results.

We define the popularity as the number of *Wikipedia* page views. Since less important events tend to contain more noise (unrelated images) in the image search results, we only consider *Events* that were viewed at least 100 times per day in average. Images of historic events also contain more noise in form of, e.g., drawings and scans, which are typically less relevant for news. To further reduce noise, we put more emphasis on significant events in the last decade(s). Thus, we used the page views v_e and the number of years a_e an *Event* e dates back to calculate the desired amount of images $|I_e|$ to crawl from *Bing*:

$$|I_e| = \min\left(1{,}000, \frac{|I_{max}| \cdot v_e}{\max(1, a_e) \cdot v_s}\right) . \quad (1)$$

In this regard, v_s denotes the number of overall views of all *Events* $\hat{E}_s$ that are a children of a *Leaf Event Node* according to the *Wikidata* knowledge graph. This normalization is used to achieve a more equal distribution of images crawled for *Leaf Event Nodes*, because some event types are less popular than others, e.g. *skeleton* compared to *basketball*. For sampling we used $|I_{max}| = 40{,}000$ and downloaded a maximum of $1{,}000$ images for the most popular events that represent the respective *Leaf Event Node*.

C. Detailed Dataset Statistics

We presented the *Visual Event Classification Dataset (VisE-D)* including two testing sets in Section 3.3 of the paper. In this section, detailed statistics on the image distribution for the *Leaf Event Nodes* are provided in Figures 1-3 (page 3-5). The illustrations also provide the complete list of *Leaf Event Nodes* used in the paper.

D. Ontologies

We provide Web-links to explore the *Ontologies* created in the paper. Different version of the *Visual Event Ontology (VisE-O)* according to the paper Sections 3.2.2 - 3.2.4 can be explored here:

- Initial ontology: link
- Disambiguated ontology: link
- Refined ontology: link

The *Ontologies* of the benchmark datasets used in Section 5.3.3 are visualized here:

- *WIDER* ontology: link
- *SocEID* ontology: link
- *RED* ontology: link

USAGE: After opening an *Ontology*, the *Leaf Event Nodes* (blue), *Branch Event Nodes* (orange), and *Root Node* (yellow) as well as their *Relations* are displayed. By clicking on a specific *Event Node* additional information such as the *Wikidata ID* and related child (*Incoming*) and parent (*Outgoing*) nodes are shown. In addition, the search bar can be used to directly access a specific *Event Node*.

Results using the probabilities $\hat{\mathbf{y}}_L^o$ for classification

	Loss	**WS**	**RR**	**Accuracy**			***JSC***	***CS***
				Top1	**Top3**	**Top5**		
C	L_c			77.4	89.8	93.6	84.7	87.7
O^{cel}	L_o			66.9	83.2	88.6	80.2	84.5
O^{cel}_{ω}	L^{cel}_o	$\omega, \omega_L = 1$		67.7	83.1	88.9	80.3	84.5
$O^{cel}_{6\omega}$	L^{cel}_o	$\omega, \omega_L = 6$		79.8	91.0	94.3	86.5	89.2
$O^{cel}_{6\omega}$+RR	L^{cel}_o	$\omega, \omega_L = 6$	✓	**81.9**	**91.7**	**94.6**	**87.9**	**90.4**
O^{cel}_{γ}	L^{cel}_o	γ		66.7	83.6	89.9	78.3	82.7
O^{cel}_{γ}+RR	L^{cel}_o	γ	✓	73.2	87.2	91.8	82.4	86.0
O^{cos}	L^{cos}_o			67.5	77.8	81.5	82.2	86.2
O^{cos}_{ω}	L^{cos}_o	$\omega, \omega_L = 1$		72.5	83.8	87.7	84.1	87.6
$O^{cos}_{6\omega}$	L^{cos}_o	$\omega, \omega_L = 6$		80.4	90.7	93.6	86.4	89.0
$O^{cos}_{6\omega}$+RR	L^{cos}_o	$\omega, \omega_L = 6$	✓	80.9	90.1	93.3	86.9	89.5
O^{cos}_{γ}	L^{cos}_o	γ		81.3	90.1	93.6	87.3	89.7
O^{cos}_{γ}+RR	L^{cos}_o	γ	✓	80.9	90.4	93.1	87.0	89.5
$CO^{cel}_{6\omega}$+RR	$L_c + L^{cel}_o$	$\omega, \omega_L = 6$	✓	81.6	**91.7**	94.5	87.5	90.0
CO^{cos}_{γ}	$L_c + L^{cos}_o$	γ		**81.9**	90.8	93.5	**87.9**	**90.4**

Table 1. Results on the manually annotated *VisE-Bing* test set using the probabilities $\hat{\mathbf{y}}_L^o$ for classification in combination with different loss functions, weighting schemes (WS), and ontology redundancy removal (RR).

Results using the probabilities $\hat{\mathbf{y}}_L^{cos}$ for classification

	Loss	**WS**	**RR**	**Accuracy**			***JSC***	***CS***
				Top1	**Top3**	**Top5**		
C	L_c			77.4	89.8	93.6	84.7	87.7
O^{cel}	L_o			68.0	77.5	81.0	82.1	86.4
O^{cel}_{ω}	L^{cel}_o	$\omega, \omega_L = 1$		68.0	78.2	82.4	81.6	85.8
$O^{cel}_{6\omega}$	L^{cel}_o	$\omega, \omega_L = 6$		79.7	89.9	92.0	86.5	89.2
$O^{cel}_{6\omega}$+RR	L^{cel}_o	$\omega, \omega_L = 6$	✓	**81.5**	90.8	92.9	**87.8**	**90.3**
O^{cel}_{γ}	L^{cel}_o	γ		66.3	80.5	85.8	78.3	82.9
O^{cel}_{γ}+RR	L^{cel}_o	γ	✓	72.7	84.7	88.2	82.3	86.0
O^{cos}	L^{cos}_o			68.8	78.8	82.1	83.9	87.7
O^{cos}_{ω}	L^{cos}_o	$\omega, \omega_L = 1$		72.5	82.9	85.2	84.7	88.1
$O^{cos}_{6\omega}$	L^{cos}_o	$\omega, \omega_L = 6$		80.1	89.5	92.0	86.3	89.0
$O^{cos}_{6\omega}$+RR	L^{cos}_o	$\omega, \omega_L = 6$	✓	80.8	89.3	91.9	86.9	89.5
O^{cos}_{γ}	L^{cos}_o	γ		79.8	87.5	89.6	86.6	89.4
O^{cos}_{γ}+RR	L^{cos}_o	γ	✓	78.3	86.6	88.3	86.1	89.0
$CO^{cel}_{6\omega}$+RR	$L_c + L^{cel}_o$	$\omega, \omega_L = 6$	✓	81.4	**91.0**	**93.1**	87.4	89.9
CO^{cos}_{γ}	$L_c + L^{cos}_o$	γ		81.4	90.5	92.5	87.2	89.8

Table 2. Results on the manually annotated *VisE-Bing* test set using the probabilities $\hat{\mathbf{y}}_L^{cos}$ for classification in combination with different loss functions, weighting schemes (WS), and ontology redundancy removal (RR).

E. Results using other Inference Strategies

We have evaluated the ontology-driven approaches using an inference strategy that combines two different probabilities $\hat{\mathbf{y}}_L = \hat{\mathbf{y}}_L^o \odot \hat{\mathbf{y}}_L^{cos}$ (Section 4.2.3). The results using the individual probabilities $\hat{\mathbf{y}}_L^o$ or $\hat{\mathbf{y}}_L^{cos}$ are provided in Table 1 and 2. In general, the probabilities $\hat{\mathbf{y}}_L^o$ provide slightly better results, in particular for the top-3 and top-5 accuracy. We believe that *Leaf Event Nodes* with shorter paths in the *Ontology* tend to achieve higher probabilities $\hat{\mathbf{y}}_L^{cos}$, as the overall (accumulated) weight of *Branch Event Nodes* is lower for the respective *Subgraph*. However, similar results are achieved in comparison to the reported numbers of the combined strategy presented in Table 2 of the paper. Thus, the results allow the same conclusion with respect to the overall performance of the ontology-driven loss functions and weighting schemes.

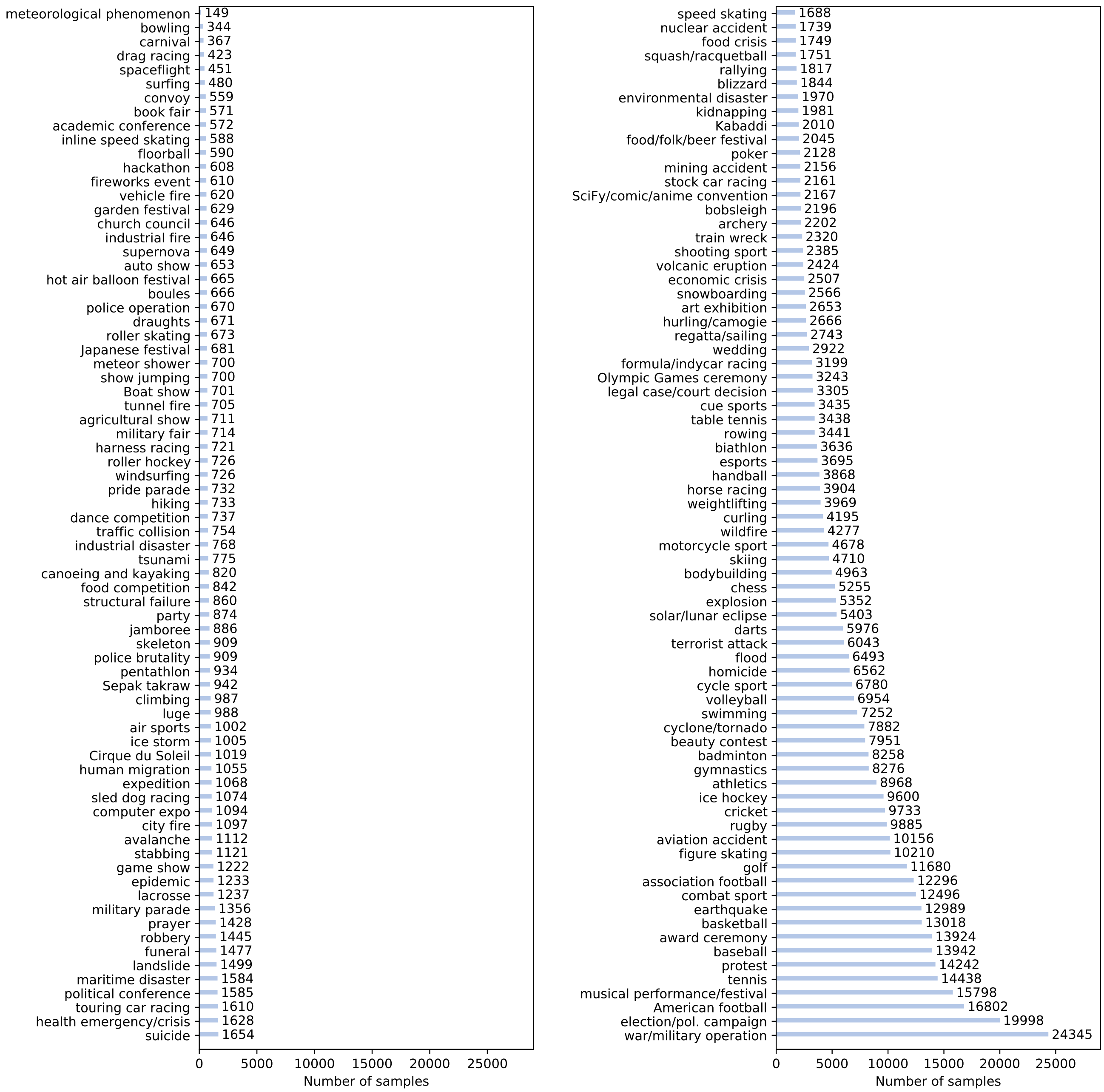

Figure 1. Number of training images for all *Leaf Event Nodes* in the *Visual Event Classification Dataset (VisE-D)*

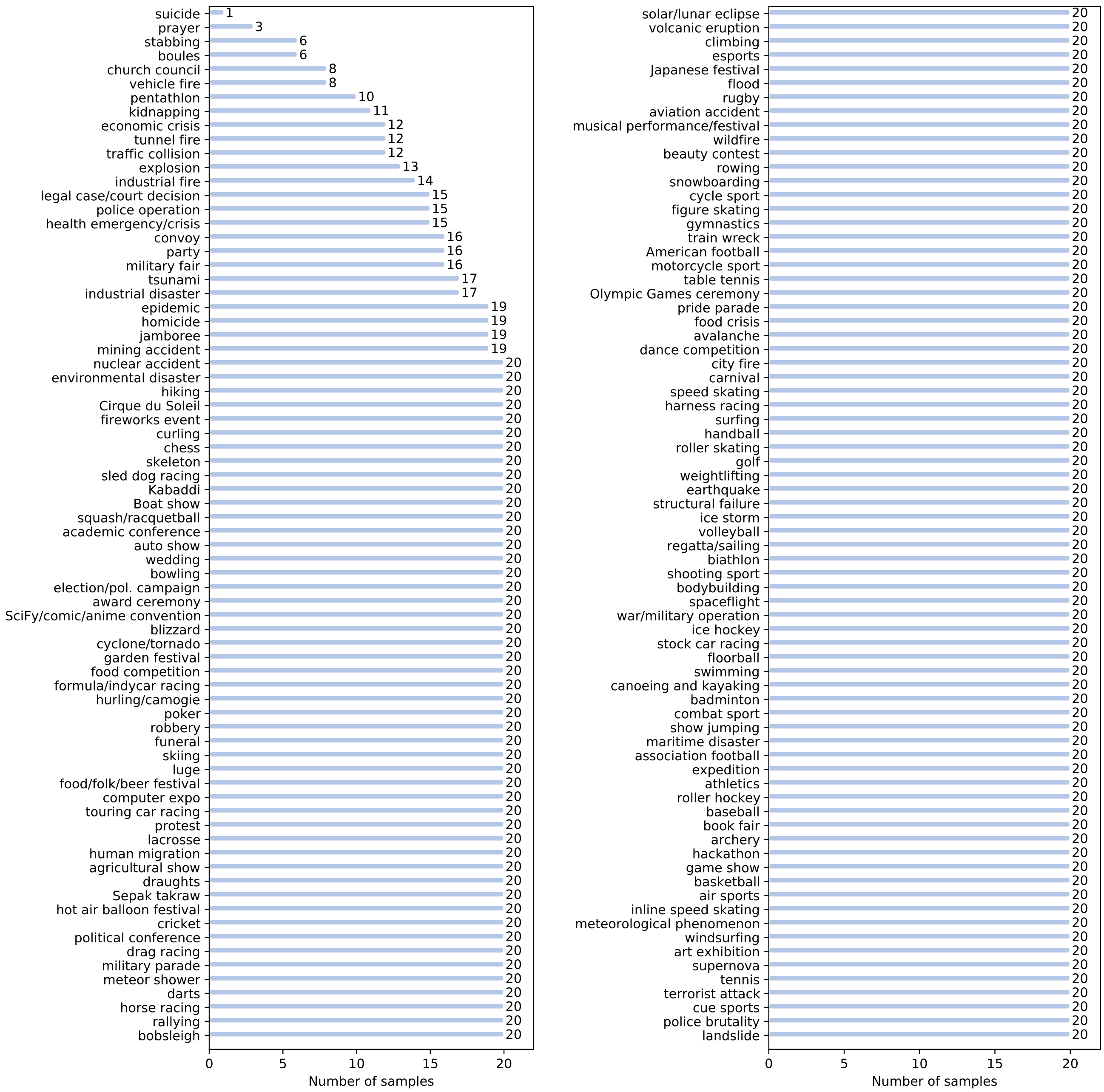

Figure 2. Number of images for all *Leaf Event Nodes* in the manually annotated *VisE-Bing* test set.

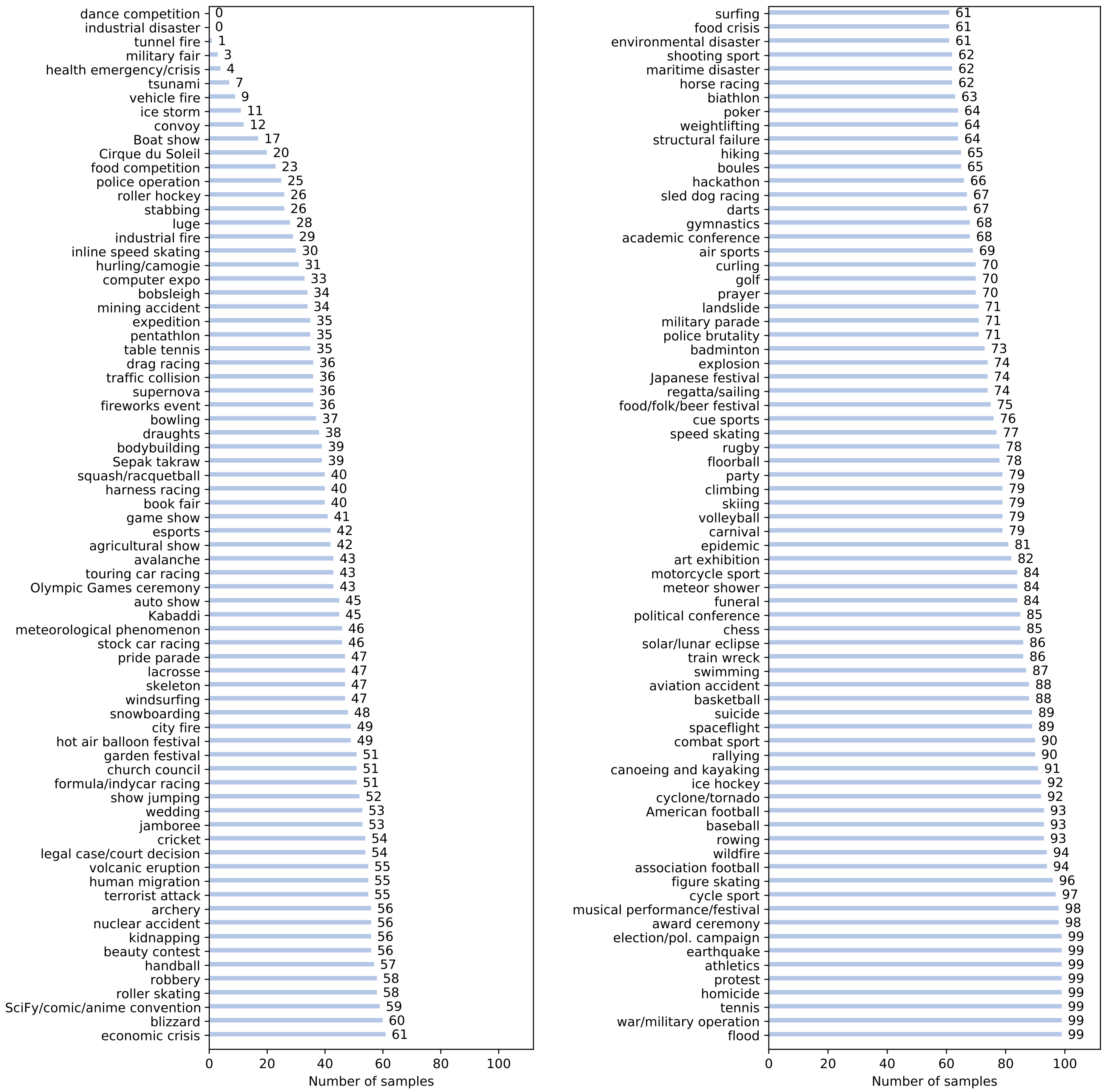

Figure 3. Number of images for all *Leaf Event Nodes* in the *VisE-Wiki* test set.