Knowledge Graph Dataset for Semantic Enrichment of Picture Description in NAPS Database

Marko Horvat [1,*], Gordan Gledec [1], Tomislav Jagušt [1] and Zoran Kalafatić [2]

1 Department of Applied Computing, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, HR-10000 Zagreb, Croatia; gordan.gledec@fer.hr (G.G.); tomislav.jagust@fer.hr (T.J.)
2 Department of Electronics, Microelectronics, Computer and Intelligent Systems, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, HR-10000 Zagreb, Croatia; zoran.kalafatic@fer.hr
* Correspondence: marko.horvat3@fer.hr

Abstract: This data description introduces a comprehensive knowledge graph (KG) dataset with detailed information about the relevant high-level semantics of visual stimuli used to induce emotional states stored in the Nencki Affective Picture System (NAPS) repository. The dataset contains 6808 systematically manually assigned annotations for 1356 NAPS pictures in 5 categories, linked to WordNet synsets and Suggested Upper Merged Ontology (SUMO) concepts presented in a tabular format. Both knowledge databases provide an extensive and supervised taxonomy glossary suitable for describing picture semantics. The annotation glossary consists of 935 WordNet and 513 SUMO entities. A description of the dataset and the specific processes used to collect, process, review, and publish the dataset as open data are also provided. This dataset is unique in that it captures complex objects, scenes, actions, and the overall context of emotional stimuli with knowledge taxonomies at a high level of quality. It provides a valuable resource for a variety of projects investigating emotion, attention, and related phenomena. In addition, researchers can use this dataset to explore the relationship between emotions and high-level semantics or to develop data-retrieval tools to generate personalized stimuli sequences. The dataset is freely available in common formats (Excel and CSV).

Dataset: https://github.com/mhorvat/NAPSKGDataset.

Dataset License: CC BY-NC-SA 4.0

Keywords: picture stimuli; affective pictures databases; image tagging; knowledge representation; affective computing

1. Summary

Stimulation of emotional states is the process of intentionally causing a person to experience a particular emotion [1]. This can be achieved through a variety of means, such as words, pictures, sounds, visualization, role-playing, or other experiential exercises [1]. The goal of stimulating emotional states is to elicit a specific emotional response from a person for use in therapy or research in psychology. For example, one such commonplace therapeutic intervention that uses visualizations and emotion-provoking images is cognitive behavioral therapy (CBT), which focuses on helping people identify and change negative thought patterns and behaviors contributing to their problems [2]. CBT can be used to treat a wide range of conditions, including anxiety, depression, and addiction [2]. In therapeutic interventions, the goal of emotion stimulation is to help people become more aware of their emotions and how they affect their thoughts and behaviors and to provide them with tools and strategies for managing their emotions more effectively [3]. This can ultimately lead to improved mental health and well-being [3,4].

One common way to elicit emotions is by using emotion-evoking images [1]. These images or pictures can be shown to people individually or in groups, and their emotional responses are commonly measured through self-reports and physiological features [5]. Emotion-evoking pictures specifically prepared for the controlled stimulation of emotions in laboratory settings are stored in affective picture databases along with their additional semantic, emotion, and context descriptors [5,6]. Based on their intended use of intentionally provoking specific emotional states, these documents are often referred to as stimuli, while pictures and videos are commonly referred to as visual stimuli [1,6].

The presented knowledge graph (KG) dataset is an extension of the Nencki Affective Picture System (NAPS) repository containing information about the relevant high-level semantics of visual stimuli [7]. The NAPS was built by the Polish Nencki Institute of Experimental Biology with the intention of providing researchers with an additional number of visual stimuli with high image quality in different categories that can be used in different areas of affective research [7]. The original database contains 1356 realistic, high-quality photographs divided into five disjointed categories: people, faces, animals, objects, and landscapes. Only parts of the photos are content-neutral because they were selected to evoke a specific emotional response in the general population. Since its introduction, the NAPS has been expanded with several additional datasets that are specialized for different domains of research in emotion processing. The extensions are: (i) NAPS Basic Emotions (NAPS BE), containing normative ratings based on the discrete model of emotions and additional dimensional ratings for a subset of 510 pictures from the original NAPS [8]; (ii) the Nencki Affective Picture System (NAPS ERO), with an additional 200 visual stimuli accompanied with self-reported subjective ratings by homosexual and heterosexual men and women ($N = 80$) of emotional valence and arousal [9]; and (iii) the Children-Rated Subset, which is the most recent extension of NAPS and includes 1128 pictures from the original NAPS database that were rated as appropriate for children based on various criteria and expert judgment [10]. The latter affective ratings were collected from a sample of $N = 266$ children aged 8–12 years [10]. One of the main important features of the NAPS set as a whole (i.e., with all its extensions) is that it combines a relatively large number of pictures with normative ratings identified according to both dimensional and categorical (discrete) emotion theories [11] compared with other stimuli sets, and it also contains additional multiword semantic descriptions organized into different topics, which are all significant features contributing to the successful construction of stimuli sequences for the elicitation of emotional reactions [7,12].

The motivation for the development of the presented dataset is found in the limitations imposed by the inadequacy and diversity of existing semantic description models used for the annotation of stimuli in contemporary affective multimedia databases. Today, these databases are described loosely and with unsupervised vocabularies; they are domain-dependent and have different models and formats. New multimedia stimuli cannot simply be added to affective multimedia databases, but require a separate set of affective ratings to be acquired through psychological experimentation with participants [1,13,14].

Recently in our previous research, we conducted a systematic online survey of domain experts in emotion stimulation and estimation. The survey results showed that researchers predominantly identify and retrieve relevant stimuli manually, which is time-consuming and labor-intensive [15]. This is due to two reasons: (1) insufficient semantic descriptors and (2) limitations of the existing stimuli retrieval software. The survey of domain experts further revealed that the quality of semantic descriptors significantly impacts user satisfaction. The findings also highlighted the importance of a user-friendly, AI-based tool for the efficient retrieval of affective pictures, particularly those that are labeled with high-quality semantic descriptors. As a result, semantically enriching current multimedia stimuli databases with additional material and KG descriptions have been shown to be a critical requirement that has the potential to dramatically improve precision, efficiency, and user satisfaction [15]. In this context, the creation of novel KG annotation datasets has emerged as a promising solution to enhance the efficiency of document retrieval and

improve the overall management of stimuli repositories. The integration of KGs in unstructured affective multimedia databases can facilitate semantic understanding and reasoning over complex data relationships, allowing for more accurate and efficient identification of relevant documents.

The remainder of this paper is structured as follows: Section 2 explains how the dataset could be employed to semantically enrich the NAPS database and contribute to the success of personalized emotion elicitation. WordNet and Suggested Upper Merged Ontology (SUMO) KGs are described with examples of their applications for the rich semantic description of NAPS pictures. In Section 3, we present the structure and format of the generated dataset, analyze the relationship between the knowledge graph concepts and the distribution of terms and concepts, and discuss the implications of our results for identifying relevant information in each of the five categories of NAPS stimuli. Section 4 provides an overview of the data collection process and the methodology for identifying knowledge graph concepts in stimuli pictures. Finally, in Section 5, we summarize our main observations and conclusions and outline plans for future research.

2. Advantages of Semantic Enrichment in Stimuli Retrieval and Personalized Emotion Stimulation

Semantic enrichment of a dataset involves adding additional information to the dataset that helps to describe and contextualize the content better [16]. The semantic enrichment process aims to make the dataset more useful for analysis, search, and information retrieval by improving its ability to be understood by humans and machines [17]. Typically, this includes adding free-text labels, tags, or differently structured metadata that give additional meaning to the data. However, semantic enrichment can be further improved by linking data from knowledge taxonomies or ontologies, either specialized for specific domains or describing general knowledge [18]. In this paper, we used the WordNet lexical knowledge graph [19,20] and general, shared, and reusable ontology SUMO [21] containing formally described concepts as the vocabularies for the labeling of entities and stimuli semantic enrichment. Our approach is explained in detail in Sections 2.1 and 2.2. An example of a thorough semantic enrichment of a NAPS visual stimulus using SUMO formal concepts is illustrated in Figure 1.

The semantic enrichment of emotion stimuli is essential for achieving higher accuracy and precision in retrieval from affective multimedia databases and, subsequently, better personalization of stimuli sequences [22]. Personalization in emotion elicitation can be defined as a process of selecting optimal stimuli for a single subject or a group of subjects who share some collective knowledge, heritage, experiences, attitudes, or perceptions that collectively determine the effect and meaning of the stimuli for the subject. The optimization criterion depends on the goals of the exposure. But regardless of the intended purpose, the stimuli must affect the subjects' cognition, behavior, and emotional states in a precise and timely manner. The desired stimulus effect and its dynamics, nature, and magnitude must be deliberately predetermined in the personalized stimuli before the exposure to ensure the expected impact on the elicitation, estimation, and regulation of emotion [1,14,23].

In the context of computerized or computer-assisted emotion elicitation exposure, personalization is effectively an interactive and often iterative process of constructing stimulus sequences as a time-dependent series of individual virtual reality (VR) or multimedia stimuli. Other stimuli such as haptic, olfactory, and vestibular stimuli [24,25] are also amenable to the personalization process, although they are less commonplace in practice, require specialized hardware, and do not have as standardized stimuli databases as the more common audiovisual stimuli. The necessary prerequisite for any personalized computer-assisted exposure is identifying the content that should have powerful significance to a specific subject. The effects of these stimuli must produce observable manifestations that an expert or specialized computer acquisition system can unambiguously identify. Clear examples of such objective phenomena are changes in physiological signals, facial

expressions, and vocal expressions. Each of them can be monitored with a number of specialized devices, such as sensors for heart rate (HR), skin temperature (SKT) and skin conductance (SC), ECG, voice and video recorders, or neuroimaging devices (fMRI, MRI, PET, EEG, MEG) [26,27]. By objectively measuring these manifestations, the success of the procedure and the personalization itself can be verified.
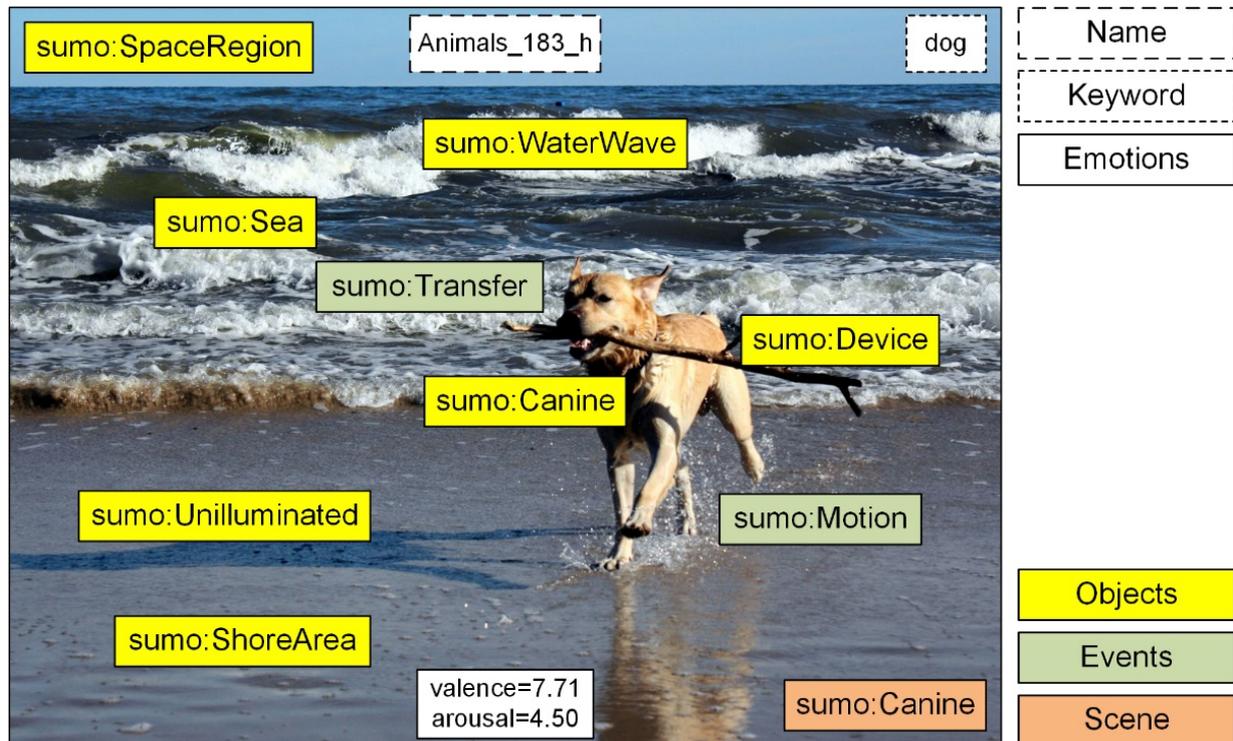

Figure 1. Illustration of a thorough semantic enrichment of a NAPS visual stimulus, Animals_183_h. The original descriptors (stimulus name, keyword, and the emotional dimensions valence and arousal) are expanded using Suggested Upper Merged Ontology (SUMO) concepts for labeling. The enrichment occurs in three stages: (1) objects within the picture are detected and labeled with KG concepts (highlighted in yellow), (2) events as actions performed by the detected objects are identified and tagged with corresponding concepts (depicted in green), (3) the global scene context is captured and labeled using an appropriate concept (represented in orange). The original NAPS metadata (white rectangles) is considerably more semantically restricted. Adapted with permission from Ref. [7], 2014, Springer.

The most important purpose of semantic enrichment in affective multimedia databases is to enable faster, simpler, and more accurate retrieval of relevant stimuli from affective multimedia databases to achieve a personalized emotion elicitation process. By doing this, semantic enrichment facilitates the creation of personalized emotion elicitation sequences. Also, as a secondary effect, the semantic enrichment process can help identify patterns and insights into the relationships between semantics and emotions that may not be apparent from the original data alone.

In the presented KG dataset, the semantic enrichment was carefully carried out manually by a group of raters. This process strictly followed a specific methodology, which is detailed in Section 4.

2.1. Representation of Stimuli Semantics with WordNet Knowledge Graph

WordNet is an extensive lexical database of the English language, developed by the Cognitive Science Laboratory at Princeton University, and is a well-known tool for describing the meanings of words and their mutual relationships [19,20]. WordNet is structured as a graph, with each word or term represented as a node in the graph. The

nodes are organized into so-called synsets ("sets of synonyms"), which represent groups of synonyms (words with similar meanings), and the relationships between words are represented as edges connecting the nodes. This allows WordNet to capture the relationships between words, such as hypernyms (more general terms), hyponyms (more specific terms), meronyms (terms part of a larger whole), etc. Hypernyms and hyponyms are usually referred to as IS-A relations and meronyms as PART-OF. In this respect, WordNet is a useful knowledge data source for a high-level description of picture content because: (1) it defines a very large and supervised labeling glossary, and (2) the labels are organized in a taxonomy as a knowledge graph [28].

A knowledge graph is a structured representation of real-world entities and the relationships between them [28]. In other words, knowledge graphs are structured representations of knowledge that model entities, attributes, and relationships between them in a graph-like structure. They enable the integration and organization of heterogeneous data sources, including textual, visual, and audio data. In this respect, WordNet's hierarchical structure of labeled concepts and its rich set of properties and relationships between concepts make it well suited to be used as a knowledge graph and to determine semantic similarities between different concepts in the graph [29].

In this approach, as illustrated in Figure 2, WordNet terms are used as tags or semantic annotations of affective pictures, and KG relations expand the descriptions.
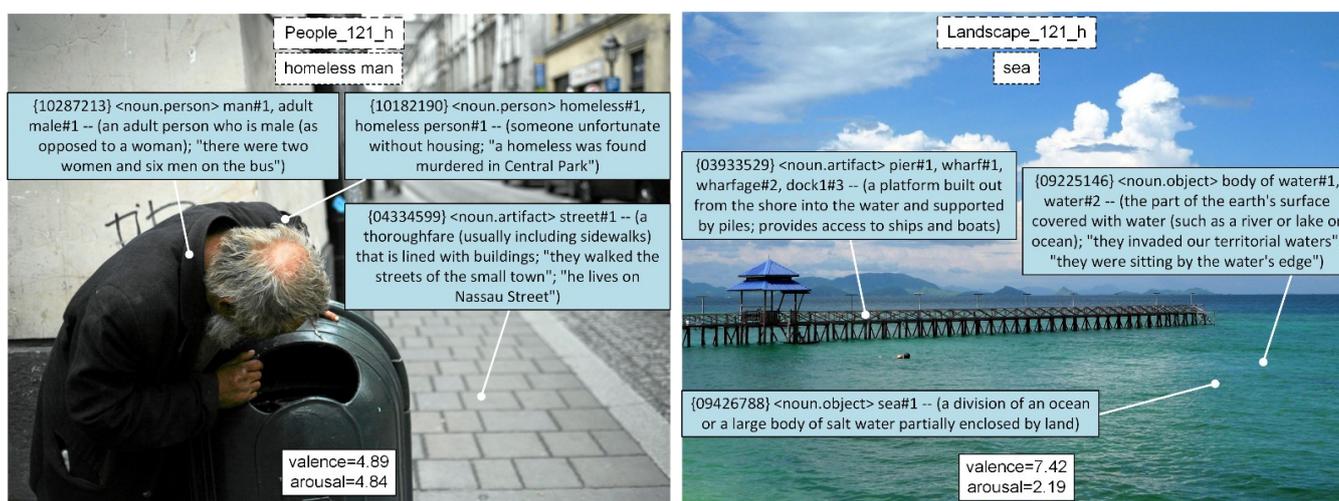

Figure 2. Semantic expansion of two NAPS stimuli, People_121_h (**left**) and Landscape_121_h (**right**), using WordNet KG entities from the presented dataset. The original labels for these stimuli were "homeless man" and "sea", respectively. The original NAPS metadata is represented by white rectangles, while the WordNet labels from the presented dataset are depicted in light blue rectangles. Adapted with permission from Ref. [7], 2014, Springer.

Building on our earlier studies [30–32], we have found that the utilization of WordNet knowledge graphs for annotating visual data is an effective strategy for enhancing the efficacy of information retrieval in multimedia stimuli databases. In our previous research, we developed a model for describing and retrieving stimuli pictures using WordNet and demonstrated the benefits of this approach using a custom software tool for this purpose [30]. The results were encouraging, and after $N = 40$ queries, showed an average precision of 68.93% and an average relevant document count of 6.15. The highest achieved precision was 84.21% for the first stimuli pictures retrieved in the results. However, the number of pictures in the experimental dataset labeled with WordNet terms needed to be higher for a thorough evaluation of retrieval performance [31,32].

2.2. Description of Relevant High-Level Visual Stimuli Semantics Using SUMO

For the formal representation of the semantics of complex stimuli, the dataset uses the general knowledge ontology SUMO to go beyond WordNet [21]. SUMO (http://www.

ontologyportal.org; accessed on 15 August 2023) is one of the most comprehensive freely available formal upper, core, and common-sense ontologies [21]. It was developed within the IEEE P1600.1 Standard Upper Ontology Working Group (SUO WG). Today, it is owned and maintained by the IEEE. Its large knowledge base contains over 25,000 terms and 80,000 axioms. The available mappings from SUMO to WordNet help express the concepts in natural language terms [33], which facilitates the extension of the framework to existing tools for the informal representation of multimedia (especially pictures) with semantic networks and lexical ontologies. In addition, SUMO is the only formal ontology mapped to the entire WordNet lexicon. Because of numerous advantageous features and comparative advantages in the formal representation of multimedia semantics over other candidate upper ontologies, we selected SUMO to develop the presented corpus. As an illustration of the high-level semantics ontology annotations, the stimulus People_172_v is originally described with the keyword "man swinging". But in the presented dataset, this is first expanded to three WordNet KG synsets, "{09225146} <noun.object> body of water#1, water#2"; "{10287213} <noun.person> man#1, adult male#1"; and "{04371774} <noun.artifact> swing#2", and then mapped to the subsuming SUMO concepts "WaterArea", "RecreationOrExerciseDevice", and "Man", as shown in Figure 3.
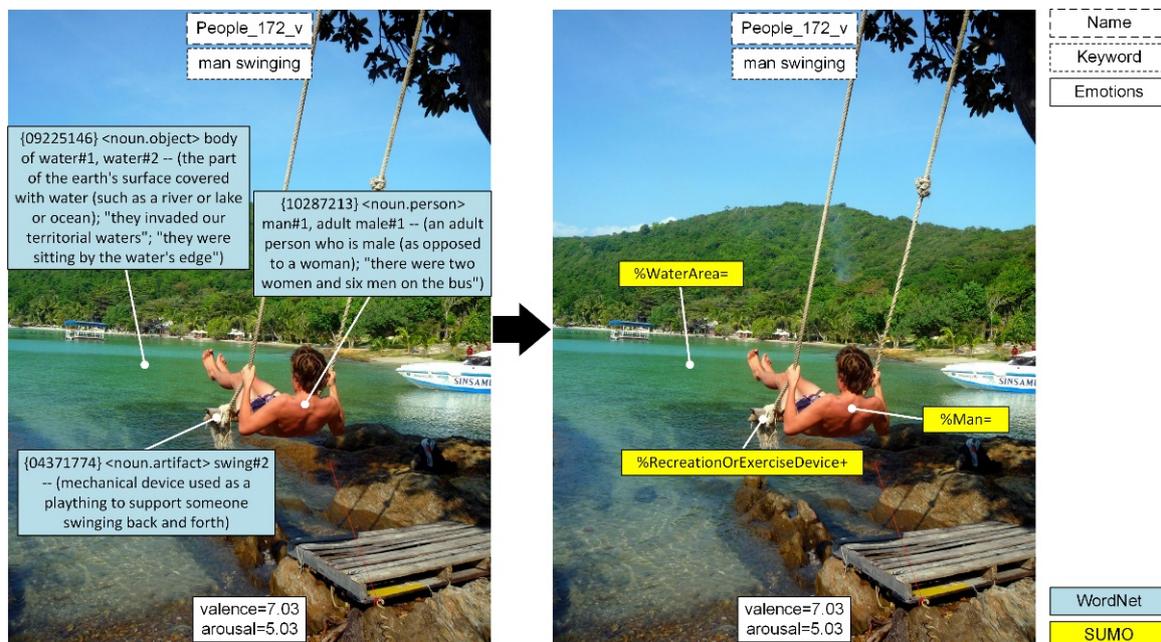

Figure 3. Suggested Upper Merged Ontology (SUMO) semantic expansion, from left to right, of the NAPS stimulus People_172_v. Objects in the picture are labeled with SUMO concepts (highlighted yellow) mapped from WordNet synsets (in light blue). The original NAPS metadata, including the picture filename, keyword, and the emotional dimensions valence and arousal, are in white rectangles. Adapted with permission from Ref. [7], 2014, Springer.

3. Data Description

The dataset for the semantic enrichment of picture descriptions in the NAPS stimuli database with KGs is represented in a structured tabular form. It is organized into rows and columns resembling a table, with each row describing one NAPS picture and each cell containing specific information for a corresponding attribute. The dataset comprises 30 comma-separated value (CSV) files and 15 Microsoft Excel (XLSX) files, for 45 files in total. The CSVs are more suitable for automated software processing and the Excel files for data examination and manual processing.

The first group of five CSVs, NAPS_WordNet_Animals.csv, NAPS_WordNet_Faces.csv, NAPS_WordNet_Landscapes.csv, NAPS_WordNet_Objects.csv, and NAPS_WordNet_People.csv,

contain the WordNet KGs associated with one of the NAPS picture categories Animals, Faces, Landscapes, Objects, and People, respectively.

Each row has the mandatory attributes or columns 'Picture_ID', 'Category', and 'Description', which are identical to the attributes in the NAPS database. The attribute 'Picture_ID' is the most important, as it represents a unique identifier for each NAPS picture (e.g., Animals_001_h, Faces_001_h, Landscapes_001_h, Objects_001_h, People_001_h). As such, it may be used for querying and integrating the KG dataset and the NAPS database. The attribute 'Category' denotes one of the five NAPS categories, and 'Description' represents the original, single free-text keyword loosely describing the picture content (e.g., "dead stork", "children with a dog", "concentration camp", "burning car", "sad woman"). In the NAPS, only the 'Description' attribute is available for descriptions of semantics. In addition to these three mandatory attributes, each row in the presented dataset contains at least one column containing WordNet KGs describing the picture. These columns are labeled 'WordNet_1', 'WordNet_2', 'WordNet_3', 'WordNet_4', 'WordNet_5', 'WordNet_6', and 'WordNet_7'.

The first group of five CSV files contains only WordNet synset IDs without any other descriptive information. These files are the most suitable for machine processing and database indexing. The first 10 rows in the NAPS_WordNet_Animals.csv datafile are provided in Table 1.

Table 1. The first 10 rows in the NAPS_WordNet_Animals.csv datafile. The attributes WordNet_n (n = 1, ..., 3) describe each NAPS picture with the WordNet KG and contain only the synset ID, for simpler processing.

Picture_ID	Category	Description	WordNet_1	WordNet_2	WordNet_3
Animals_001_h	Animals	dead stork	00095280	02002075	
Animals_002_v	Animals	lion	02129165		
Animals_003_h	Animals	snake	01726692		
Animals_004_v	Animals	wolf	02114100	01045719	
Animals_005_h	Animals	bat	02139199		
Animals_006_v	Animals	snake	01726692		
Animals_007_h	Animals	wolf	02114100	15043763	
Animals_008_v	Animals	fighting chickens	01660444	07644967	01792158
Animals_009_v	Animals	cat	02121620	08438533	
Animals_010_h	Animals	sick kitten	02541302	02122948	

The second group of five CSV files have the suffix "_Complete". Their filenames are NAPS_WordNet_Animals_Complete.csv, NAPS_WordNet_Faces_Complete.csv, NAPS_WordNet_Landscapes_Complete.csv, NAPS_WordNet_Objects_Complete.csv, and NAPS_WordNet_People_Complete.csv. These files have the same structure as the first group, and they also describe the NAPS stimuli with WordNet KGs. But the CSV files from this group contain the entire descriptive content of the WordNet synsets, including their ID, term type, enumerated synonyms, and other information. As an example, the first three rows of the NAPS_WordNet_Complete_Animals.csv are shown in Table 2.

The third group of five CSV files in the presented dataset contain ontology concepts describing NAPS stimuli using the formal vocabulary defined by SUMO. They are: NAPS_SUMO_Animals.csv, NAPS_SUMO_Faces.csv, NAPS_SUMO_Landscapes.csv, NAPS_SUMO_Objects.csv, and NAPS_SUMO_People.csv. Each row in these documents also describes the semantics of a single picture and has the same tabular structure as the WordNet CSV files. The first three columns are 'Picture_ID', 'Category', and 'Description', while the remaining seven columns are denoted 'SUMO_1', 'SUMO_2', 'SUMO_3', 'SUMO_4', 'SUMO_5', 'SUMO_6', and 'SUMO_7'. Table 3 shows a sample of the dataset in NAPS_SUMO_Animals.csv.

Table 2. The first three rows in the NAPS_WordNet_Complete_Animals.csv datafile.

Picture_ID	Category	Description	WordNet_1	WordNet_2
Animals_001_h	Animals	dead stork	{00095280} <adj.all> dead1#1 (vs. alive1#1) -- (no longer having or seeming to have or expecting to have life; "the nerve is dead"; "a dead pallor"; "he was marked as a dead man by the assassin")	{02002075} <noun.animal> stork#1 -- (large mostly Old World wading birds typically having white-and-black plumage)
Animals_002_v	Animals	lion	{02129165} <noun.animal> lion#1, king of beasts#1, Panthera leo#1 -- (large gregarious predatory feline of Africa and India having a tawny coat with a shaggy mane in the male)	
Animals_003_h	Animals	snake	{01726692} <noun.animal> snake#1, serpent#1, ophidian#1 -- (limbless scaly elongate reptile; some are venomous)	

Table 3. The first 10 rows in the NAPS_SUMO_Animals.csv datafile. Each attribute SUMO_n (n = 1, ..., 3) describes the semantics of a NAPS picture with a single SUMO concept.

Picture_ID	Category	Description	SUMO_1	SUMO_2	SUMO_3
Animals_001_h	Animals	dead stork	%Dead=	%Bird+	
Animals_002_v	Animals	lion	%Lion=		
Animals_003_h	Animals	snake	%Snake=		
Animals_004_v	Animals	wolf	%Canine+	%RadiatingSound+	
Animals_005_h	Animals	bat	%Mammal+		
Animals_006_v	Animals	snake	%Snake=		
Animals_007_h	Animals	wolf	%Canine+	%Snowing=	
Animals_008_v	Animals	fighting chickens	%ViolentContest+	%ChickenMeat+	%Rooster=
Animals_009_v	Animals	cat	%Feline+	%Forest=	
Animals_010_h	Animals	sick kitten	%DiseaseOrSyndrome+	%Kitten=	

The structure of the 15 Excel files (XLSX) is identical to those of the already described CSV files. Each row and column (i.e., attribute) in the Excel files corresponds to a row and column in the CSV files, ensuring consistency between the two file types. This enables seamless data comparison and processing as well as consistent data usage and analysis.

The CSV files were created using Microsoft Excel, which uses the semicolon (;) character as the default column separator. To facilitate interoperability with all data processing tools, the dataset contains an additional 15 CSV files with the comma (,) as the separator. These additional CSV files are denoted with the suffix "_CommaDelimited" in their filenames.

Because of the tabular structure of the knowledge graph dataset, when exporting empty or null values in certain cells to a CSV format, the data are displayed as consecutive separators (e.g., ";;", ";;;", ";;;;", etc.). However, standard spreadsheets and text editors can handle such data.

It is important to mention that the presented dataset is licensed under "Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)". Other parties are free to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material for noncommercial purposes. If other parties remix, transform, or build upon the presented dataset, they must distribute the dataset under the same license as the original. Other parties must give appropriate credit, provide a link to this license, and indicate if changes were made.

3.1. Data Utilization

The presented KG dataset incorporates WordNet and SUMO vocabularies, providing a more sophisticated representation of the real-world entities of the NAPS affective pictures' content. This semantics annotation method is more structured and expressive than the

traditional free-text keyword model. As a result, the process of querying, retrieving, and analyzing affective pictures becomes much more streamlined and efficient.

However, the dataset does not contain affective pictures and emotion information. To utilize the dataset for document retrieval, it is necessary to request the NAPS repository for nonprofit academic research purposes from the Nencki Institute of Experimental Biology, Laboratory of Brain Imaging (LOBI), at https://lobi.nencki.gov.pl/research/8/ (accessed on 15 August 2023).

An example of querying would involve utilizing the attribute 'Picture_ID' from both the KG dataset and the NAPS database as the common link or key. When aiming for document retrieval from the NAPS database, one would begin by selecting a specific 'Picture_ID' from the KG dataset. This 'Picture_ID' would then be matched with the corresponding 'Picture_ID' in the NAPS database. By ensuring that both attributes match, one can perform a join operation to merge the relevant data from the two records. As a result of this join operation based on the common attribute 'Picture_ID', the user will retrieve comprehensive document details from the NAPS database enriched with the semantic information from the KG. This method facilitates precise and detailed document searching and utilizes the content of both datasets: high-level semantics from the KG dataset and pictures and emotions from the NAPS database.

The integration of the KG dataset, with separate WordNet and SUMO data tables, and the NAPS data table is illustrated in Figure 4. The KG WordNet dataset is represented on the left side of the figure. It is illustrated as a table with multiple columns labeled as 'Picture_ID', 'Category', 'Description', 'WordNet_1', 'WordNet_2', 'WordNet_3', 'WordNet_4', 'WordNet_5', 'WordNet_6', and 'WordNet_7'. These labels indicate the various columns or attributes found within this dataset. Similarly, the KG SUMO dataset is represented on the right side of the figure. The NAPS database table in the middle consists of three columns: 'Picture_ID', 'Category', and 'Description'. The attribute 'Picture_ID' serves as the primary key (PK) for all data tables and also as the foreign key for joining the KG WordNet table (FK1) and KG SUMO table (FK2) with 1:1 relationship cardinality (i.e., a multiplicity relationship attribute).
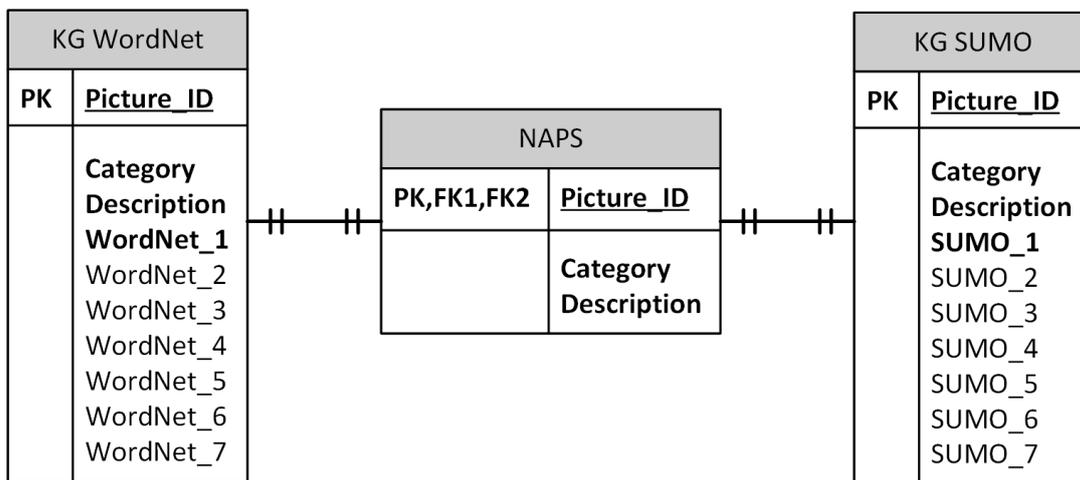

Figure 4. Integration of the KG dataset (WordNet and SUMO) with the NAPS database can be accomplished by utilizing the 'Picture_ID' attribute as the foreign key with 1:1 relationship cardinality. This approach semantically enriches the original NAPS attribute 'Description' with 14 KG attributes.

Effectively, by using the presented dataset, the NAPS attribute 'Description' is semantically enriched with the 14 KG attributes 'WordNet_1'–'WordNet_7' and 'SUMO_1'–'SUMO_7'.

3.2. Data Distribution

It is important to analyze the distribution of data points in the corpus while analyzing the attributes of the generated dataset. The presented dataset comprises 6808 systematically manually assigned annotations or labels for 1356 NAPS pictures in 5 categories. Out of 6808 labels, 3429 are WordNet concepts and 3379 are SUMO concepts. This glossary comprises 935 unique WordNet synsets and 513 SUMO concepts. Because of the higher abstraction, substantially fewer SUMO concepts are needed to semantically describe pictures. Figure 5 depicts the frequency distribution of the synsets and ontology concepts, highlighting the dataset's most and least commonly utilized KG entities. This provides insight into the overall distribution of the KGs and their usage trends.
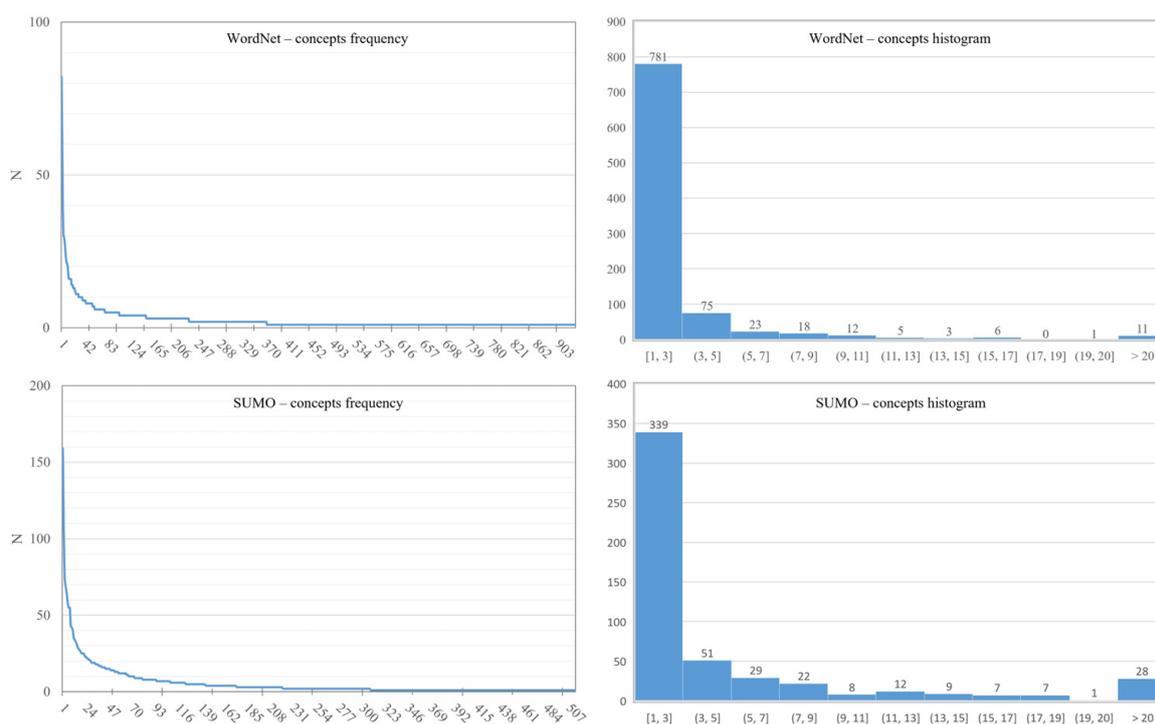
Figure 5. Frequency distribution of WordNet (**upper row**) and SUMO (**lower row**) entities in the dataset. The diagrams show the frequency count (*N*) on the *y*-axis and the entity ordinal on the *x*-axis. On the right, the histograms group the entities into bins with a width of 2. In total, the manual annotation of the NAPS pictures involved 935 different WordNet synsets and 513 SUMO concepts.

The 10 most used annotating synsets in the dataset, with their respective frequencies in brackets, are: "{10287213} <noun.person> man#1, adult male#1" (82); "{10787470} <noun.person> woman#1, adult female#1" (61); "{08436759} <noun.group> vegetation#1, flora#1, botany#1" (39); "{06878071} <noun.communication> smile#1, smiling#1, grin#1, grinning#1" (30); "{13104059} <noun.plant> tree#1" (30); "{03544360} <noun.artifact> house#1" (29); "{02084071} <noun.animal> dog#1, domestic dog#1, Canis familiaris#1" (27); "{05600637} <noun.body> face#1, human face1#1" (24); "{09436708} <noun.object> sky#1" (22); and "{02121620} <noun.animal> cat#1, true cat#1" (21). Likewise, the 10 most frequent SUMO concepts are: "Man" (159), "Woman" (107), "HumanChild" (74), "Smiling" (69), "Device" (64), "WaterArea" (58), "Human" (55), "SubjectiveAssessmentAttribute" (55), "BotanicalTree" (43), and "Plant" (42).

Another important data feature is how many KGs are used to describe each NAPS picture. The distribution remains consistent across all the common descriptive statistical parameters and the five picture categories, as shown in Figure 6. The thorough adherence to the tight formal rules utilized in the picture annotation process can be credited to the uniform distribution. The approach enables reliable and accurate labeling, contributing to the general consistency of the KG distribution throughout the dataset.
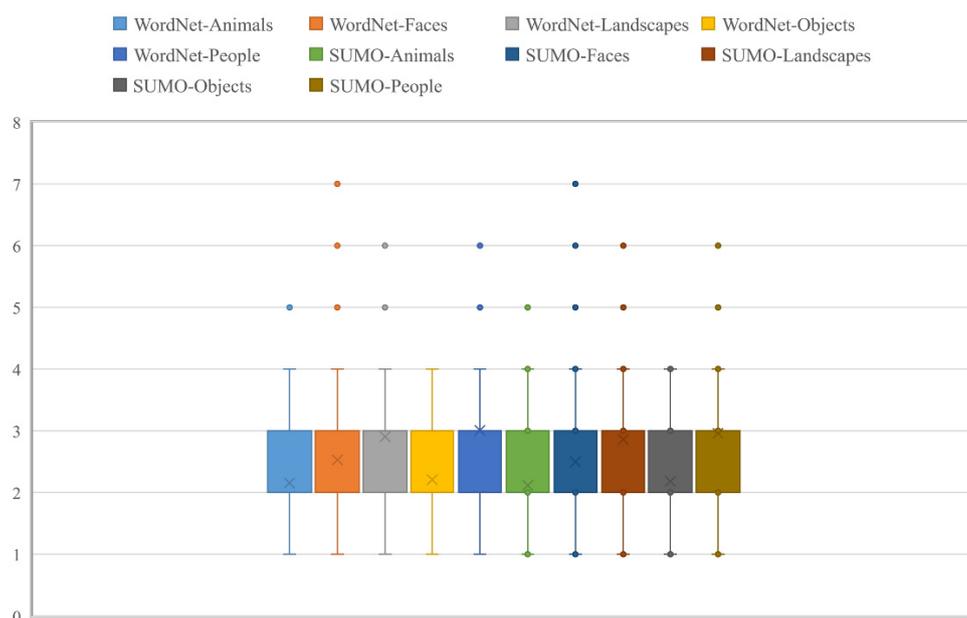
Figure 6. Box-and-whisker diagrams showing the distribution of WordNet and SUMO entities for description of each picture in five categories: Animals, Faces, Landscapes, Objects, and People.

The numbers of WordNet synsets and SUMO concepts used to semantically describe each picture are very similar in the presented dataset. This can be attributed to the usage of mappings, which effectively connect each synset to a related concept. However, it is important to consider that the overall number of distinct ontology concepts is much lower than the entire number of unique synsets. This is primarily due to the fact that within the ontology, multiple synsets have been allocated to identical or subsuming concepts. This mapping technique contributes to the simplification and consolidation of the semantic representation of pictures, resulting in a more compact and coherent ontology-based description.

4. Methods

A formal protocol was established and followed by a group of $N = 15$ raters or annotators to ensure accurate and consistent labeling of the NAPS pictures. The raters were senior computer engineering students, and the group leader was a university professor. All had previous experience in labeling various multimodal documents and web pages. The protocol called for each member of the group to observe all pictures in a random sequence and independently label each picture based on a predefined set of criteria. The raters were instructed to detect and label the entities in a picture using the following annotation methodology: (1) detect objects in the scene (obligatory), (2) determine adjectives pertaining to the detected objects (if possible), (3) identify verbs and adverbs describing an action depicted in the scene (if possible), and (4) describe the whole scene (if possible). A quality control process was implemented to ensure that the labeling process was consistent and accurate over time. This included regular checks by the team leader, and the members could ask for additional training sessions to reinforce the criteria and the labeling process. If any discrepancies or inconsistencies were found, they were documented and discussed by the group before a consensus was reached on the final labeling of each picture. Once all pictures had been labeled, the team leader reviewed each picture and verified the accuracy of the labeling.

In addition, all team members were trained on the criteria and labeling process prior to the labeling process to ensure consistency and accuracy. A set of pictures retrieved from the Internet were used for the training to ensure that members were not familiar with the NAPS pictures before they began the labeling process. The entire annotation methodology employed to label the NAPS pictures in the presented knowledge graph dataset is described with the UML activity diagram in Figure 7.
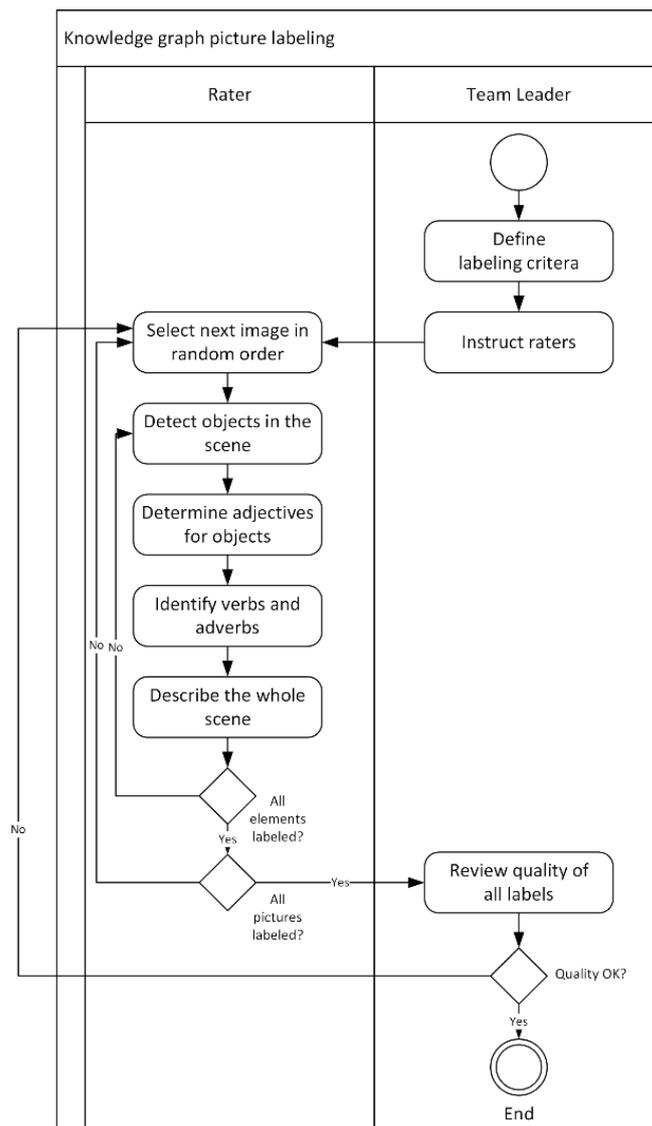
Figure 7. UML activity diagram describing the methodology employed by a group of raters and the team leader to systematically label NAPS pictures in the presented knowledge graph dataset.

When multiple raters are involved in the labeling process, it is possible that each individual may interpret the criteria differently, resulting in labeling inconsistencies. However, a formal protocol ensures that all labeling team members use the same criteria and interpret the instructions in the same way. This protocol ensures that the labeling process is thorough and that the resulting dataset is consistent, accurate, and reliable, which is crucial for future dataset applications.

5. Summary

The choice of relevant stimuli is frequently limited in the available emotionally annotated databases that store different types of stimuli. Much can and should be done to improve the functionality and interoperability of existing emotionally annotated databases. The semantic enrichment of multimedia databases is a crucial step toward enhancing their accessibility and usability. By incorporating higher-level semantic metadata, such as knowledge graphs with formal concepts and relationships, we can facilitate more efficient identification and retrieval of relevant content from large multimedia databases. This improves the user experience and supports a wide range of applications, such as multimedia retrieval, stimuli recommendation, and emotion elicitation personalization.

The dataset uses the WordNet knowledge graph, the SUMO upper ontology, and SUMO to WordNet mappings to provide rich, high-level semantic expressivity with interfaces to commonly used models and existing systems. The dataset improves the knowledge reuse, interoperability, and formalization of picture stimuli information over current methods for representing stimuli based on keywords or tags. All these features enable formal, consistent, and systematic annotation of affective multimedia content and document properties.

Future research should explore innovative approaches and tools for semantic enrichment, with a focus on addressing the challenges of scalability, semantic heterogeneity, and accuracy. In this regard, the knowledge graph dataset should be expanded to include other affective multimedia databases researchers use most frequently in addition to NAPS. By transforming keywords to high-level concepts and mapping them to an upper core formal ontology, it will be possible to achieve semantic integration of different multimedia stimuli databases; i.e., to combine emotion-elicitation documents from various sources, formats, or systems to allow for meaningful interpretation and analysis. In our previous research, we created the first versions of such ontologies [34,35] and plan to expand their knowledge models further and use them in the continuation of our work.

The presented dataset could be used to explore the finer relationships between emotions and semantics in affective multimedia in general (e.g., the semantic gap), and especially those encountered in specific stimuli sequences for certain domains. Both could provide further insights into the affective data properties and move to a more overarching affective model that includes emotion, cognition, behavior, and action properties.

In addition, the presented dataset could be used as a foundation to develop novel data retrieval software tools using emotional and semantic descriptors, allowing for a more efficient construction of personalized emotion-elicitation sequences for therapeutic interventions, personalized education, and interactive entertainment. For example, these tools could be used by therapists to help patients with anxiety or depression find images that evoke positive emotions. It could also be used by educators to create personalized learning materials that are tailored to the cognitive and emotional needs of students. Novel tools could be used to create personalized gaming experiences by tailoring the game content to the player's individual preferences and emotional state. Such intelligent tools would also enable meaningful and accurate data analysis in the domains of pedagogy, education, psychology, neuroscience, and cognitive sciences.

Finally, recent advancements in artificial intelligence (AI), such as machine learning (ML), deep learning (DL), and natural language processing (NLP), have the potential to significantly impact the development of KGs for affective multimedia databases. For example, DL with NLP techniques can be used to automatically detect objects in images or videos and extract semantic information from text descriptions. This information can then be used to create new semantic descriptors for affective multimedia or to improve the accuracy of existing descriptors. ML techniques can also be used to learn the relationships between semantic descriptors and emotional responses. This information can then be used to develop additional semantic and emotion descriptor datasets or more effective data retrieval tools for personalized stimuli sequences.

References

1. Coan, J.A.; Allen, J.J. (Eds.) *Handbook of Emotion Elicitation and Assessment*; Oxford University Press: Oxford, UK, 2007.
2. Kazantzis, N.; Luong, H.K.; Usatoff, A.S.; Impala, T.; Yew, R.Y.; Hofmann, S.G. The processes of cognitive behavioral therapy: A review of meta-analyses. *Cogn. Ther. Res.* **2018**, *42*, 349–357. [CrossRef]
3. Wilhelm, S.; Weingarden, H.; Ladis, I.; Braddick, V.; Shin, J.; Jacobson, N.C. Cognitive-behavioral therapy in the digital age: Presidential address. *Behav. Ther.* **2020**, *51*, 1–14. [CrossRef]
4. Montana, J.I.; Matamala-Gomez, M.; Maisto, M.; Mavrodiev, P.A.; Cavalera, C.M.; Diana, B.; Mantovani, F.; Realdon, O. The Benefits of emotion Regulation Interventions in Virtual Reality for the Improvement of Wellbeing in Adults and Older Adults: A Systematic Review. *J. Clin. Med.* **2020**, *9*, 500. [CrossRef] [PubMed]
5. Wang, Y.; Song, W.; Tao, W.; Liotta, A.; Yang, D.; Li, X.; Gao, S.; Sun, Y.; Ge, W.; Zhang, W.; et al. A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances. *Inf. Fusion* **2022**, *83*, 19–52. [CrossRef]
6. Horvat, M. A Brief Overview of Affective Multimedia Databases. In *Central European Conference on Information and Intelligent Systems*; Faculty of Organization and Informatics: Varaždin, Croatia, 2017; pp. 3–9.
7. Marchewka, A.; Żurawski, Ł.; Jednorog, K.; Grabowska, A. The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behav. Res. Methods* **2014**, *46*, 596–610. [CrossRef]
8. Riegel, M.; Żurawski, Ł.; Wierzba, M.; Moslehi, A.; Klocek, Ł.; Horvat, M.; Grabowska, A.; Michałowski, J.; Marchewka, A. Characterization of the Nencki Affective Picture System by discrete emotional categories (NAPS BE). *Behav. Res. Methods* **2016**, *48*, 600–612. [CrossRef] [PubMed]
9. Wierzba, M.; Riegel, M.; Pucz, A.; Leśniewska, Z.; Dragan, W.Ł.; Gola, M.; Jednorog, K.; Marchewka, A. Erotic subset for the Nencki Affective Picture System (NAPS ERO): Cross-sexual comparison study. *Front. Psychol.* **2015**, *6*, 1336. [CrossRef]
10. Zamora, E.V.; Richard's, M.M.; Introzzi, I.; Aydmune, Y.; Urquijo, S.; Olmos, J.G.; Marchewka, A. The Nencki Affective Picture System (NAPS): A Children-Rated Subset. *Trends Psychol.* **2020**, *28*, 477–493. [CrossRef]
11. Horvat, M.; Stojanović, A.; Kovačević, Ž. An overview of common emotion models in computer systems. In Proceedings of the 45th Jubilee International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2022), Opatija, Croatia, 23–27 May 2022; pp. 1152–1157. [CrossRef]
12. Horvat, M.; Jović, A.; Burnik, K. Investigation of Relationships between Discrete and Dimensional Emotion Models in Affective Picture Databases Using Unsupervised Machine Learning. *Appl. Sci.* **2022**, *12*, 7864. [CrossRef]
13. Uhrig, M.K.; Trautmann, N.; Baumgärtner, U.; Treede, R.D.; Henrich, F.; Hiller, W.; Marschall, S. Emotion elicitation: A comparison of pictures and films. *Front. Psychol.* **2016**, *7*, 180. [CrossRef]
14. Blanco-Ruiz, M.; Sainz-de-Baranda, C.; Gutiérrez-Martín, L.; Romero-Perales, E.; López-Ongil, C. Emotion elicitation under audiovisual stimuli reception: Should artificial intelligence consider the gender perspective? *Int. J. Environ. Res. Public Health* **2020**, *17*, 8534. [CrossRef] [PubMed]
15. Horvat, M.; Jerčić, P. A Survey on Usage of Multimedia Databases for Emotion Elicitation: A Quantitative Report on How Content Diversity Can Improve Performance. In Proceedings of the 2023 46th MIPRO ICT and Electronics Convention (MIPRO 2023), Opatija, Croatia, 22–26 May 2023; pp. 1148–1154. [CrossRef]
16. Abgaz, Y.; Rocha Souza, R.; Methuku, J.; Koch, G.; Dorn, A. A Methodology for Semantic Enrichment of Cultural Heritage Images Using Artificial Intelligence Technologies. *J. Imaging* **2021**, *7*, 121. [CrossRef] [PubMed]
17. Silvello, G.; Bordea, G.; Ferro, N.; Buitelaar, P.; Bogers, T. Semantic Representation and Enrichment of Information Retrieval Experimental Data. *Int. J. Digit. Libr.* **2017**, *18*, 145–172. [CrossRef]
18. Simeone, D.; Cursi, S.; Acierno, M. BIM Semantic-Enrichment for Built Heritage Representation. *Autom. Constr.* **2019**, *97*, 122–137. [CrossRef]
19. Fellbaum, C. *WordNet*; MIT Press: Cambridge, MA, USA, 1998.
20. Miller, G.A. WordNet: A lexical database for English. *Commun. ACM* **1995**, *38*, 39–41. [CrossRef]
21. Pease, A.; Niles, I.; Li, J. The suggested upper merged ontology: A large ontology for the semantic web and its applications. In Proceedings of the Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web, Edmonton, AB, Canada, 28 July 2002; Volume 28.
22. Doré, B.P.; Silvers, J.A.; Ochsner, K.N. Toward a personalized science of emotion regulation. *Soc. Personal. Psychol. Compass* **2016**, *10*, 171–187. [CrossRef]

23. Ćosić, K.; Popović, S.; Horvat, M.; Kukolja, D.; Dropuljić, B.; Kovač, B.; Jakovljević, M. Computer-aided psychotherapy based on multimodal elicitation, estimation and regulation of emotion. *Psychiatr. Danub.* **2013**, *25*, 340–346.

24. Radianti, J.; Majchrzak, T.A.; Fromm, J.; Wohlgenannt, I. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Comput. Educ.* **2020**, *147*, 103778. [CrossRef]

25. Yeomans, J.S.; Li, L.; Scott, B.W.; Frankland, P.W. Tactile, acoustic and vestibular systems sum to elicit the startle reflex. *Neurosci. Biobehav. Rev.* **2002**, *26*, 1–11. [CrossRef]

26. Dzedzickis, A.; Kaklauskas, A.; Bucinskas, V. Human emotion recognition: Review of sensors and methods. *Sensors* **2020**, *20*, 592. [CrossRef]

27. Egger, M.; Ley, M.; Hanke, S. Emotion recognition from physiological signal analysis: A review. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 35–55. [CrossRef]

28. Fensel, D.; Şimşek, U.; Angele, K.; Huaman, E.; Kärle, E.; Panasiuk, O.; Toma, I.; Umbrich, J.; Wahler, A. Introduction: What is a knowledge graph? In *Knowledge Graphs: Methodology, Tools and Selected Use Cases*; Springer: Cham, Switzerland, 2020; pp. 1–10. [CrossRef]

29. Zhu, G.; Iglesias, C.A. Computing semantic similarity of concepts in knowledge graphs. *IEEE Trans. Knowl. Data Eng.* **2016**, *29*, 72–85. [CrossRef]

30. Horvat, M.; Grbin, A.; Gledec, G. Labeling and retrieval of emotionally-annotated images using WordNet. *Int. J. Knowl.-Based Intell. Eng. Syst.* **2013**, *17*, 157–166. [CrossRef]

31. Horvat, M.; Grbin, A.; Gledec, G. WNtags: A Web-Based Tool For Image Labeling And Retrieval With Lexical Ontologies. *Front. Artif. Intell. Appl.* **2012**, *243*, 585–594. [CrossRef]

32. Horvat, M.; Vuković, M.; Car, Ž. Evaluation of keyword search in affective multimedia databases. In *Transactions on Computational Collective Intelligence XXI: Special Issue on Keyword Search and Big Data*; Lecture Notes in Computer Science, 9630; Springer: Berlin/Heidelberg, Germany, 2016; pp. 50–68. [CrossRef]

33. Niles, I.; Pease, A. Linking lexicons and ontologies: Mapping wordnet to the suggested upper merged ontology. In Proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE 03), Las Vegas, NV, USA, 23–26 June 2003; pp. 23–26.

34. Horvat, M.; Bogunović, N.; Ćosić, K. STIMONT: A Core Ontology for Multimedia Stimuli Description. *Multimed. Tools Appl.* **2014**, *73*, 1103–1127. [CrossRef]

35. Horvat, M. StimSeqOnt: An ontology for formal description of multimedia stimuli sequences. In Proceedings of the 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 28 September–2 October 2020; pp. 1134–1139. [CrossRef]