

Tipos de publicaciones y métricas

Prof. Rosa Paccotacya Yanque

rypaccotacya@ucsp.edu.pe

E. P. de Ciencia de la Computación - UCSP

Types of publications

- Thesis
- Research paper
- Review
- Survey

Others:

- Extended abstract
- Poster

Thesis

- Document containing the research and findings that students submit to get the professional qualification or degree.
- In a thesis, the researcher puts forth his/her conclusion. The researcher also gives evidence in support of the conclusion.
- The thesis contains a **broader description of the subject matter**. In contrast, the research paper contains a **narrow description** of the subject matter.

Survey

- According to the definition of survey paper provided by IEEE Communications Surveys & Tutorials journal, "The term survey, as applied here, is defined to mean a survey of the literature. A survey article should provide a comprehensive review of developments in a selected area".
- In ACM Computing Survey, survey paper is described as "A paper that summarizes and organizes recent research results in a novel way that integrates and adds understanding to work in the field. A survey article emphasizes the classification of the existing literature, developing a perspective on the area, and evaluating trends."
- In Elsevier journal of Computer Science Review, you will see here that "Critical review of the relevant literature" is required a component of every typical survey paper.

Engineering Applications of Artificial Intelligence 110 (2022) 104743





Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Survey paper

A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects



Absalom E. Ezugwu ^{a,*}, Abiodun M. Ikotun ^a, Olaide O. Oyelade ^a, Laith Abualigah ^{b,c,**}, Jeffery O. Agushaka ^a, Christopher I. Eke ^d, Andronicus A. Akinyelu ^e

- a School of Computer Science, University of KwaZulu-Natal, Pietermaritzburg Campus, Pietermaritzburg 3201, South Africa
- ^b Faculty of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan
- School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, 11800, Malaysia
- ^d Department of Computer Science, Faculty of Natural Science, Federal University of Lafia, Lafia, Nigeria
- Department of Computer Science and Informatics, University of the Free State, 9301 Bloemfontein, South Africa

ARTICLE INFO

Keywords: Clustering Clustering algorithms, partitioning Data mining Hierarchical clustering

Automatic clustering K-Means Optimization algorithms, Machine learning Unsupervised learning Supervised learning

ABSTRACT

Clustering is an essential tool in data mining research and applications. It is the subject of active research in many fields of study, such as computer science, data science, statistics, pattern recognition, artificial intelligence, and machine learning. Several clustering techniques have been proposed and implemented, and most of them successfully find excellent quality or optimal clustering results in the domains mentioned earlier. However, there has been a gradual shift in the choice of clustering methods among domain experts and practitioners alike, which is precipitated by the fact that most traditional clustering algorithms still depend on the number of clusters provided a priori. These conventional clustering algorithms cannot effectively handle real-world data clustering analysis problems where the number of clusters in data objects cannot be easily identified. Also, they cannot effectively manage problems where the ontimal number of clusters for a high-dimensional dataset cannot be easily determined. Therefore, there is a need for improved, flexible, and efficient clustering techniques, Recently, a variety of efficient clustering algorithms have been proposed in the literature, and these algorithms produced good results when evaluated on real-world clustering problems. This study presents an up-to-date systematic and comprehensive review of traditional and state-of-the-art clustering techniques for different domains. This survey considers clustering from a more practical perspective. It shows the outstanding role of clustering in various disciplines, such as education, marketing, medicine, biology, and bioinformatics. It also discusses the application of clustering to different fields attracting intensive efforts among the scientific community, such as big data, artificial intelligence, and robotics. This survey paper will he heneficial for both practitioners and researchers. It will serve as a good reference point for researchers and practitioners to design improved and efficient state-of-the-art clustering algorithms.

1. Introduction

Clustering (an aspect of data mining) is considered an active method of grouping data into many collections or clusters according to the similarities of data points features and characteristics (zian, 2010; Abauligab, 2019). Over the past years, dozens of data clustering techniques have been proposed and implemented to solve data clustering tep-niques have been proposed and implemented to solve data clustering problems (Zhou et al., 2019; Abauligab et al., 2018a, bl. In general, clustering analysis techniques can be divided into two main groups: hierarchical and partitional (Ten., 2018). Although methods in these

two groups have proved to be very effective and efficient, they generally depend on providing prior knowledge or information of the exact number of clusters for each dataset to be clustered and analyzed (Chang et al., 2010). More so, when dealing with real-world datasets, it is normal not to expect or have any prior information regarding the number of naturally occurring groups in the data objects (Liu et al., 2011). Therefore, the concept of automatic data clustering algorithms is introduced to address this limitation. Automatic clustering algorithms refer to any clustering techniques used to automatically determine

https://doi.org/10.1016/j.engappai.2022.104743

Received 7 December 2021; Received in revised form 16 January 2022; Accepted 3 February 2022 Available online 23 February 2022

0952-1976/© 2022 Elsevier Ltd. All rights reserved.



Explainable Reinforcement Learning: A Survey

Erika Puiutta^(⊠) and Eric M. S. P. Veith

OFFIS - Institute for Information Technology, Escherweg 2, 26121 Oldenburg, Germany {erika.puiutta,eric.veith}@offis.de

Abstract. Explainable Artificial Intelligence (XAI), i.e., the development of more transparent and interpretable AI models, has gained increased traction over the last few years. This is due to the fact that, in conjunction with their growth into powerful and ubiquitous tools, AI models exhibit one detrimental characteristic; a performance-transparency trade-off. This describes the fact that the more complex a model's inner workings, the less clear it is how its predictions or decisions were achieved. But, especially considering Machine Learning (ML) methods like Reinforcement Learning (RL) where the system learns autonomously. the necessity to understand the underlying reasoning for their decisions becomes apparent. Since, to the best of our knowledge, there exists no single work offering an overview of Explainable Reinforcement Learning (XRL) methods, this survey attempts to address this gap. We give a short summary of the problem, a definition of important terms, and offer a classification and assessment of current XRL methods. We found that a) the majority of XRL methods function by mimicking and simplifying a complex model instead of designing an inherently simple one, and b) XRL (and XAI) methods often neglect to consider the human side of the equation, not taking into account research from related fields like psychology or philosophy. Thus, an interdisciplinary effort is needed to adapt the generated explanations to a (non-expert) human user in order to effectively progress in the field of XRL and XAI in general.

Keywords: Machine learning · Explainable · Reinforcement Learning · Human-computer interaction · Interpretable

1 Introduction

Over the past decades, AI has become ubiquitous in many areas of our everyday lives. Especially Machine Learning (ML) as one branch of AI has numerous fields of application, be it transportation [57], advertisement [46], or medicine [38]. Unfortunately, the more powerful and flexible those models are, the more opaque they become, essentially making them black boxes (see Fig. 1). This trade-off is

^{*} Correspondence to: School of Mathematics, Statistics, and Computer Science, University of KwaZulu-Natal, King Edward Road, Pietermaritzburg, KwaZulu-Natal 3201, South Africa.

^{**} Corresponding author at: Faculty of Computer Sciences and Informatics, Amman Arab University, Amman 11953, Jordan.

E-mail addresse: Erugwua@ukzn.ac.za (A.E. Ezugwu), 220078470@stu.ukzn.ac.za (A.M. Ikotun), oyeladeo@ukzn.ac.za (O.O. Oyelade), laythdysha@aau.edu.jo (L. Abualigah), 218088307@stu.ukzn.ac.za (J.O. Agushaka), eke.fleanyi@fulafia.edu.ng (C.I. Eke), AkinyeluAA@ufs.ac.za (A.A. Akinvelu.)

[©] IFIP International Federation for Information Processing 2020 Published by Springer Nature Switzerland AG 2020 A. Holzinger et al. (Eds.): CD-MAKE 2020, LNCS 12279, pp. 77–95, 2020. https://doi.org/10.1007/978-3-030-57321-8_5

Review

Involves the collection of all the relevant literature on a topic. However, unlike a survey article, it additionally discusses the metadata from the surveyed literature to technically compare different studies, draws conclusions on their weaknesses and strengths, and proposes future directions.

Trustworthy Artificial Intelligence: A Review

DAVINDER KAUR, SULEYMAN USLU, KALEY J. RITTICHIER, and ARJAN DURRESI, Indiana University-Purdue University Indianapolis

Artificial intelligence (AI) and algorithmic decision making are having a profound impact on our daily lives. These systems are vastly used in different high-stakes applications like healthcare, business, government, education, and justice, moving us toward a more algorithmic society. However, despite so many advantages of these systems, they sometimes directly or indirectly cause harm to the users and society. Therefore, it has become essential to make these systems safe, reliable, and trustworthy. Several requirements, such as fairness, explaniability, accountability, reliability, and acceptance, have been proposed in this direction to make these systems trustworthy. This survey analyzes all of these different requirements through the lens of the literature. It provides an overview of different approaches that can help mitigate AI risks and increase trust and acceptance of the systems by utilizing the users and society. It also discusses existing strategies for validating and verifying these systems and the current standardization efforts for trustworthy AI. Finally, we present a holistic view of the recent advancements in trustworthy AI to help the interested researchers grasp the crucial facets of the topic efficiently and offer possible future research directions.

CCS Concepts: • Computing methodologies → Cross-validation; Intelligent agents;

Additional Key Words and Phrases: Artificial intelligence, machine learning, black-box problem, trustworthy AI, explainable AI, fairness, explainability, accountability, privacy, acceptance

ACM Reference format:

Davinder Kaur, Suleyman Usłu, Kaley J. Rittichier, and Arjan Durresi. 2022. Trustworthy Artificial Intelligence: A Review. ACM Comput. Surv. 55, 2, Article 39 (January 2022), 38 pages. https://doi.org/10.1145/3491209

1 INTRODUCTION

Artificial intelligence (AI) and algorithmic decision making are transforming our lives. In to-day's world, most of our day-to-day tasks are either done or guided by machines or algorithms. This area of algorithmic decision making is not new. We have been using machines for decision making for a long time. However, today these systems have become very efficient and complex because of the availability of vast data, advanced algorithms, and high computing power. It has

This work was partially supported by the National Science Foundation (NSF) under grant 1547411 and by the U.S. Department of Agriculture (USDA), National Institute of Food and Agriculture (NIFA) (award 2017-67003-26057) via an interagency partnership between USDA-NIFA and the NSF on the research program Innovations at the Nexus of Food, Energy, and Water Systems.

Authors' address: D. Kaur, S. Uslu, K. J. Rittichier, and A. Durresi, Indiana University-Purdue University Indianapolis, Computer & Information Science, 723 W Michigan St, Indianapolis, IN 46202; emails: {davikaur, suslu, krittich}@iu.edu, adurresi@iupui.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

0360-0300/2022/01-ART39 \$15.00

https://doi.org/10.1145/3491209

Systematic Review

A systematic review is more exhaustive than a literature review as it includes both published and unpublished literature. Further, a systematic review sets out to address a question and does so by a **thorough review of existing literature** (based on an appropriate **search methodology** for a systematic search) and robust data extraction and analysis techniques.

Characteristics of publicly available skin cancer image datasets: a systematic review



David Wen, Saad M Khan, Antonio Ji Xu, Hussein Ibrahim, Luke Smith, Jose Caballero, Luis Zepeda, Carlos de Blas Perez, Alastair K Denniston, Xiaoxuan Liu*, Rubeta N Matin*

Publicly available skin image datasets are increasingly used to develop machine learning algorithms for skin cancer diagnosis. However, the total number of datasets and their respective content is currently unclear. This systematic review aimed to identify and evaluate all publicly available skin image datasets used for skin cancer diagnosis by exploring their characteristics, data access requirements, and associated image metadata. A combined MEDLINE, Google, and Google Dataset search identified 21 open access datasets containing 106 950 skin lesion images, 17 open access atlases, eight regulated access datasets, and three regulated access atlases. Images and accompanying data from open access datasets were evaluated by two independent reviewers. Among the 14 datasets that reported country of origin, most (11 [79%]) originated from Europe, North America, and Oceania exclusively. Most datasets (19 [91%]) contained dermoscopic images or macroscopic photographs only. Clinical information was available regarding age for 81662 images (76.4%), sex for 82848 (77.5%), and body site for 79561 (74.4%). Subject ethnicity data were available for 1415 images (1.3%), and Fitzpatrick skin type data for 2236 (2.1%). There was limited and variable reporting of characteristics and metadata among datasets, with substantial under-representation of darker skin types. This is the first systematic review to characterise publicly available skin image datasets, highlighting limited applicability to real-life clinical settings and restricted population representation, precluding generalisability. Quality standards for characteristics and metadata reporting for skin image datasets are needed.

Introduction

Digital health innovation has the potential to improve health care by increasing access to specialist expertise. 1.2

Collaboration (ISIC) archive.13 are increasingly used to develop machine learning algorithms for skin cancer diagnosis.24-26 Additionally, although primarily aimed for



Lancet Digit Health 2022:

Published Online November 9, 2021 https://doi.org/10.1016/ 52589-7500(21)00252-1

*Ioint last authors

Oxford University Clinical Academic Graduate School. University of Oxford, Oxford, UK (DWen BMBCh): Institute of Clinical Sciences, University of Birmingham, Birmingham, UK (DWen): Royal Berkshire Hospital, Royal Berkshire NHS Foundation Trust, Reading, UK (DWen, SM Khan MBChB): Department of Dermatology. Churchill Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK (A li Xu BMBCh. R N Matin PhD): University Hospitals

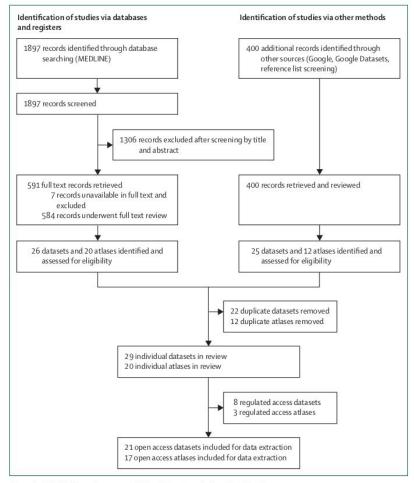


Figure 1: PRISMA flow diagram outlining dataset and atlas identification

Scene Graph Prediction with Limited Labels

Chen, Paroma Varma, Ranjay Krishna, Michael Bernstein, Christopher Ré, Li Fei-Fei Stanford University

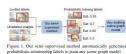
ncentsc, paroma, ranjaykrishna, msb, chrismre, feifeili}@cs.stanford.edu

ledge bases such as Visual Genome power ications in computer vision, including visual ring and captioning, but suffer from sparse. tionships. All scene graph models to date raining on a small set of visual relationships ands of training labels each. Hiring human xpensive, and using textual knowledge base thods are incompatible with visual data. In introduce a semi-supervised method that asstic relationship labels to a large number of es using few labeled examples. We analyze hips to suggest two types of image-agnostic used to generate noisy heuristics, whose outated using a factor graph-based generative s few as 10 labeled examples per relationative model creates enough training data to at our method outperforms all baseline apene graph prediction by 5.16 recall@100

S. In our limited label setting, we define a

ric for relationships that serves as an indi-778) for conditions under which our method transfer learning, the de-facto approach for mited labels.

to formalize a structured representation for Genome [27] defined scene graphs, a forlar to those widely used to represent knowl-[18, 56]. Scene graphs encode objects (e.g. e) as nodes connected via pairwise relationding) as edges. This formalization has led art models in image cantioning [3] image], visual question answering [24], relation-[26] and image generation [23]. However, ne graph models ignore more than 98% of tegories that do not have sufficient labeled Figure 2) and instead focus on modeling the



probabilistic relationship labels to train any scene graph model. few relationships that have thousands of labels [31, 49, 54].

Hiring more human workers is an ineffective solution to labeling relationships because image annotation is so tedious that seemingly obvious labels are left unannotated. To complement human annotators, traditional text-based knowledge completion tasks have leveraged numerous semi-supervised or distant supervision approaches [6, 7, 17, 34]. These methods find syntactical or lexical patterns from a small labeled set to extract missing relationships from a large unlabeled set. In text, pattern-based methods are successful, as relationships in text are usually document-agnostic (e.g. <Tokyo is capital of - Japan>). Visual relationships are often incidental: they depend on the contents of the particular image they appear in. Therefore, methods that rely on external knowledge or on patterns over concepts (e.g. most instances of dog next to frisbee are playing with it) do not generalize well. The inability to utilize the progress in text-based methods necessitates specialized methods for visual knowledge

In this paper, we automatically generate missing rela-

tionships labels using a small, labeled dataset and use these

relationships has been unexplored. While image-agnostic

generated labels to train downstream scene graph models (see Figure 1). We begin by exploring how to define imageagnostic features for relationships so they follow patterns across images. For example, ear usually consists of one object consuming another object smaller than itself, whereas look often consists of common objects: phone, laptop, or windraw (see Figure 3). These rules are not dependent on raw pixel values; they can be derived from image-agnostic features like object categories and relative spatial positions between objects in a relationship. While such rules are simmethod's improvements over transfer learning (Section 5.3). ple, their capacity to provide supervision for unannotated

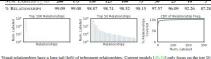


Figure 2. Visual relationships have a long tail (left) of infrequent relationships. Current models [49,54] only focus on the top 50 relationships (middle) in the Visual Genome dataset, which all have thousands of labeled instances. This ignores more than 98% of the relationships with

features can characterize some visual relationships very well. they might fail to canture complex relationships with high Textual knowledge bases were originally hand-curated by

variance. To quantify the efficacy of our image-agnostic features, we define "subtypes" that measure spatial and cateeorical complexity (Section 3) Based on our analysis, we propose a semi-supervised approach that leverages image-agnostic features to label missine relationships using as few as 10 labeled instances of each

relationship. We learn simple heuristics over these features and assign probabilistic labels to the unlabeled images using a generative model [30,46]. We evaluate our method's labeling efficacy using the completely-labeled VRD dataset [31] and find that it achieves an F1 score of 57.66, which is 11.84 points higher than other standard semi-supervised methods like label propagation [57]. To demonstrate the utility of our generated labels, we train a state-of-the-art scene graph model [54] (see Figure 6) and modify its loss function to support probabilistic labels. Our approach achieves 47.53 recall@1001 for predicate classification on Visual Genome, improving over the same model trained using only labeled instances by 40.97 points. For scene graph detection, our approach achieves within 8.65 recall@100 of the same model trained on the original Visual Genome dataset with 108× more labeled data. We end by comparing our approach to transfer learning, the de-facto choice for learning from limited labels. We find that our approach improves by 5.16 recall@100 for predicate classification, especially for relationships with high complexity, as it generalizes well to

Our contributions are three-fold. (1) We introduce the first method to complete visual knowledge bases by finding missing visual relationships (Section 5.1). (2) We show the utility of our generated labels in training existing scene graph prediction models (Section 5.2). (3) We introduce a metric to characterize the complexity of visual relationships and show it is a strong indicator ($R^2 = 0.778$) for our semi-supervised

Recall@K is a standard measure for scene graph prediction [31].

Figure 7. (a) Heuristics based on spatial features help predict <man - fly - kite>. (b) Our model learns that look is highly correlated

with phone. (c) We overfit to the importance of chair as a categorical feature for sit, and fail to identify hand as the correct relationship.

(d) We overfit to the spatial positioning associated with ride, where objects are typically longer and directly underneath the subject (e)

Given our image-agnostic features, we produce a reasonable label for class-cover-face
However, our model is incorrect, as two

typically different predicates (sit and cover) share a semantic meaning in the context of <qlasses - ? - face>.

experts to structure facts [4,5,44] (e.g. < Tokyo - capital of - Japan>). To scale dataset curation efforts, recent approaches mine knowledge from the web [9] or hire nonexpert annotators to manually curate knowledge [5, 47]. In semi-supervised solutions, a small amount of labeled text is used to extract and exploit patterns in unlabeled sentences [2, 21, 33-35, 371. Unfortunately, such approaches cannot be directly applied to visual relationships; textual relations can often be captured by external knowledge or patterns, while visual relationships are often local to an image

Visual relationships have been studied as spatial priors [14, 161, co-occurrences [51], language statistics [28,31,53], and within entity contexts [29]. Scene graph prediction models have dealt with the difficulty of learning from incomplete knowledge, as recent methods utilize statistical motife (54) or object-relationship dependencies (30, 49, 50, 55) All these methods limit their inference to the top 50 most frequently occurring predicate categories and ignore those without enough labeled examples (Figure 2)

The de-facto solution for limited label problems is transfer learning [15, 52], which requires that the source domain used for pre-training follows a similar distribution as the target domain. In our setting, the source domain is a dataset of frequently-labeled relationships with thousands of examples [30, 49, 50, 55], and the target domain is a set of limited. label relationships. Despite similar objects in source and target domains, we find that transfer learning has difficulty generalizing to new relationships. Our method does not rely on availability of a larger, labeled set of relationships; instead, we use a small labeled set to annotate the unlabeled

To address the issue of gathering enough training labels for machine learning models, data programming has emerged as a popular paradigm. This approach learns to model imperfect labeling sources in order to assign training labels to unlabeled data. Imperfect labeling sources can come from crowdsourcing [10], user-defined heuristics [8, 43], multi-instance learning [22, 40], and distant su-

cification (DDDDCT C) which are note around touth bounding



that our semi-supervised method outnerforms transfer learn-

ing which has soon more data. Eurthormore we countify



Figure 3. Relationships, such as £1y, oat, and sit can be characterized effectively by their categorical (s and o refer to subject and object, respectively) or spatial features. Some relationships like fly rely heavily only on a few features - kitos are often seen high up in the sky.

pervision [12, 32]. Often, these imperfect labeling sources to label relationships with limited data? Previous literature take advantage of domain expertise from the user. In our case, imperfect labeling sources are automatically generated

Table 2. Results for scene graph prediction tasks with n = 10 labeled examples per predicate, reported as recall@K. A state-of-the-art scene graph model trained on labels from our method outperforms those trained with labels generated by other baselines, like transfer learning.

0.01

11 10 11 08

0.00 0.01

14.02 14.51

10.98 11.28

20.83 21.44

0.00

9.01 11.01 11.64

10.16 10.84

11 99 14 40

12.58

6.74 6.83 9.67 0.01 0.07

3.20

label to every pair of object proposals. 3. Analyzing visual relationships

We define the formal terminology used in the rest of the paper and introduce the image-agnostic features that our semi-supervised method relies on. Then, we seek quantitative insights into how visual relationships can be described by the properties between its objects. We ask (1) what imageagnostic features can characterize visual relationships? and (2) given limited labels, how well do our chosen features characterize the complexity of relationships? With these in mind, we motivate our model design to generate heuristics that do not overfit to the small amount of labeled data and assign accurate labels to the larger unlabeled set

heuristics, which we aggregate to assign a final probabilistic

BASELINE In = 101

TRANSFER LEARNING

I AREL PROPAGATION [ST]

OURS (CATEG. + SPAT. + DEEP)

DECISION TREE [38]

FREO+OVERLAP

OURS (DEEP)

OURS (SPAT)

OURS (CATEG.)

A scene graph is a multi-graph G that consists of objects o as nodes and relationships r as edges. Each object o; = $\{b_i, c_i\}$ consists of a bounding box b_i and its category $c_i \in$ C where C is the set of all possible object categories (e.g. dog, frisbee). Relationships are denoted < subject. predicate - object> or <o - p - d>. p ∈ P is a predicate, such as ride and eat. We assume that we have a small labeled set $\{(o, p, o') \in D_v\}$ of annotated relationships for each predicate n. Usually these datasets are on the order of a 10 examples or fewer. For our semisupervised approach, we also assume that there exists a large set of images Dr. without any labeled relationships.

3.2. Defining image-agnostic features

It has become common in computer vision to utilize pretrained convolutional neural networks to extract features that represent objects and visual relationships [31, 49, 50]. Models trained with these features have proven robust in the presence of enough training labels but tend to overfit when presented with limited data (Section 5). Consequently, an open question arises: what other features can we utilize

has combined deep learning features with extra information extracted from categorical object labels and relative spatial object locations [25, 31]. We define categorical features, < o, -, o' >, as a concatenation of one-hot vectors of the subject o and object o'. We define spatial features as:



where b = [y, x, h, w] and b' = [y', x', h', w'] are the topleft bounding box coordinates and their widths and heights. To explore how well spatial and categorical features can describe different visual relationships, we train a simple decision tree model for each relationship. We plot the importances for the top 4 spatial and categorical features in Figure 3. Relationships like fly place high importance on the difference in y-coordinate between the subject and object, capturing a characteristic spatial pattern. Look, on the other hand, depends on the category of the objects (e.g. phone, laptop, window) and not on any spatial orientations.

To understand the efficacy of image-agnostic features.

3.3. Complexity of relationships

we'd like to measure how well they can characterize the complexity of particular visual relationships. As seen in Figure 4, a visual relationship can be defined by a number of image-agnostic features (e.g. a person can ride a bike, or a dog can ride a surfboard). To systematically define this notion of complexity, we identify subtypes for each visual relationship. Each subtype captures one way that a relationship manifests in the dataset. For example, in Figure 4, ride contains one categorical subtype with <person - ride hike > and another with < dog - ride - surfingard > Similarly, a person might carry an object in different relative spatial orientations (e.g. on her head, to her side). As shown in Figure 5, visual relationships might have significantly different degrees of spatial and categorical complexity, and therefore a different number of subtypes for each. To compute spatial subtypes, we perform mean shift clustering [11] over the spatial features extracted from all the

> 10.92 20.98 20.98

14 57

21.57

5 30

20 30 20 90 22 21

39 69 41 65 42 37

31 75 33 02 33 35

26.23 27.10 27.26



associated with a relationship.

With access to 10 or fewer labeled instances for these visual relationships, it is impossible to capture all the subtypes for given relationship and therefore difficult to learn a good representation for the relationship as a whole. Consequently, we turn to the rules extracted from image-agnostic features and use them to assign labels to the unlabeled data in order to canture a larger proportion of subtypes in each visual relationship. We posit that this will be advantageous over methods that only use the small labeled set to train a scene graph prediction model, especially for relationships with high complexity, or a large number of subtypes. In Section 5.3, we find a correlation between our definition of complexity and the performance of our method.

We aim to automatically generate labels for missing visual relationships that can be then used to train any downstream scene graph prediction model. We assume that in the longtail of infrequent relationships, we have a small labeled set $\{(o, p, o') \in D_n\}$ of annotated relationships for each predicate p (often, on the order of a 10 examples or less). As discussed in Section 3, we want to leverage image-agnostic features to learn rules that annotate unlabeled relationships.

Our approach assigns probabilistic labels to a set D_U of un-annotated images in three steps: (1) we extract image agnostic features from the objects in the labeled D. and





Figure 5. A subset of visual relationships with different levels of complexity as defined by spatial and categorica we show how this measure is a good indicator of our semi-supervised method's effectiveness compared to baselin

relationships in Visual Genome. To compute the categorical Algorithm 1 Semi-supervised Alg. subtypes, we count the number of unique object categories INPUT: {(o, p, o') ∈ D_p}∀p ∈ P — A am with multi-class labels for predicates.
 INPUT: {(o, o')} ∈ D_U} — A large untal

jects but no relationship labels.

3. INPUT: $f(\cdot, \cdot) = A$ function that entracts fea.

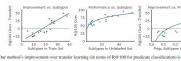
4. INPUT: $DT(\cdot) = A$ decision tree.

5. INPUT: $G(\cdot) = A$ generative model that as multiple labels for each datapoint INPUT: train(-) — Function used to train a s

Extract features and labels, X_y , $Y_y := \{f(X_U := \{(f(o,o') \text{ for } (o,o') \in D_U\}\}$ Generate heuristics by fitting J decision trees Assign labels to $(o,o') \in D_U$, $\Lambda = DT_{yo}$,

 Learn generative model G(A) and assign prob
 Train scene graph model, SGM := train(D_p.
 OUTPUT: SGM(-) from the object proposals extracted u

detector [19] on unlabeled D_U , (2) over the image-agnostic features, as factor-graph based generative mode sign probabilistic labels to the unlabe These probabilistic labels, along wit any scene graph prediction model. We in Algorithm 1 and show the end-to-e Feature extraction: Our approach u features defined in Section 3, which re box and category labels. The featur ground truth objects in D. or from o in Dr. by running existing object dete Heuristic generation: We fit decir beled relationships' spatial and cates ture image-agnostic rules that define



subtypes in the train set (left), the number of subtypes in the unlabeled set (middle), and the proportion of subtype

We also achieve within 8.65 recall@100 of ORACLE for we hypothesized earlier. TRANSFER SGDET. We generate higher quality training labels than cases when the labeled set only capti DECISION TREE and LARGE PROPAGATION leading to an the relationship's subtypes. This tr plains how OURS (CATEG + SPAT given a small portion of labeled subty

In Figure 8 (right

ince as the number

g that we approach

TEER I harts perfor-

ded examples.

13.83 and 22.12 recall@100 increase for PREDCLS. Effect of labeled and unlabeled data, In Figure 8 (left and PREDCLS per

er of labeled exam-6 Conclusion 5. 10. We observe FARNING as n de-H@100 PREDCTS observations from ed examples gives out a larger propor-

knowledge bases like Visual Genor

We introduce the first method

visual relationships. We define cate tures as image-apposite features and based generative model that uses the probabilistic labels to unlabeled im performs baselines in F1 score when tionships in the complete VRD datas be used to train scene graph predict modifications to their loss function labels. We outperform transfer learni and come close to oracle performar

Research Papers

rules are threshold-based conditions that are efined by the decision tree. To limit the comheuristics and thereby prevent overfitting, we ision trees [38] with different restrictions on feature set to produce J different decision redict labels for the unlabeled set using these bucing a $\Lambda \in \mathbb{R}^{J \times |D_U|}$ matrix of predictions

lationship (e.g., carry), we use image-agnostic features to automatically create heuristics and then use a generative model

e only use these heuristics when they have about their label: we modify A by converting abel with confidence less than a threshold osen to be 2× random) to an abstain, or no nt. An example of a heuristic is shown in he subject is above the object, it assigns a

or the predicate carry. del: These heuristics, individually, are noisy sign labels to all object pairs in Dr. As a gate the labels from all J heuristics. To do so, actor graph-based generative model popular eak supervision techniques [1, 39, 41, 45, 48]

 $L_{\theta} = \mathbb{E}_{Y \sim \pi} \left[\log \left(1 + \exp(-\theta^T V^T Y) \right) \right]$

where θ is the learned parameters π is the distribution learned by the generative model, Y is the true label, and V are features extracted by any scene graph prediction model.

relationships or scene graphs. Each scene graph contains objects localized as bounding boxes in the image along with nairwise relationships connecting them, categorized as action (e.g., carry), possessive (e.g., wear), spatial (e.g.,

features in D_U before propagating labels from D_u to D_U .

We compare to a strong frequency baselines: (FREQ) uses the object counts as priors to make relationship predictions, and FREQ+OVERLAP increments such counts only if the

labeled relationship instances. We also compare to ORACLE, which is trained with 108× more labeled data

bines all three, and OURS (CATEG. + SPAT. + WORDVEC) Figure 7(b), we correctly label look because phone i includes word vectors as richer representations of the cate- an important categorical feature. In some difficult cases,

spatial features, (CATEG. + SPAT. + DEEP) combines comobjects that have a large difference in y-coordinate. I

There are many genres

Even within areas, there exist many different genres of paper. Each genre is typically built around the claim you are making, and implies a structure to the sections and to the writing. For example:

We solve a problem: articulate the problem, explain what causes that problem and what others have done to deal with it, detail your approach, and prove that you make progress on the problem

We measure an outcome: explain that nobody has bothered understanding how a phenomenon behaves, explain how to create a study that sheds light, and report the outcomes of it

We introduce a technique: articulate a problem as above, but focus the narrative on the technique you've created, since it will generalize

Adapted from Stanford CS197

Genres imply structure

Common "We Solve A Problem" structure:

Introduction: overview and thesis

Related Work: situate your contribution relative to prior research

Approach: describe your approach and important implementation details

Evaluation: test whether your approach succeeds at its stated goals

Method

Results

Discussion: reflect on limitations, implications, and future work

Conclusion: summarize and restate your contribution

But, this will vary

Adapted from Stanford CS197 17

Contribution Type

What type of contribution is your project making?

HCI: interaction technique, system, etc.

Al: theory, system / architecture, etc.

Different contributions may have different paper structures or expectations for each paper section.

Example: HCI (Interaction Technique)

Introduction

RW

System

User Study & Results

Discussion

Limitations & Future Work

Conclusion

In-Depth Mouse: Integrating Desktop Mouse into Virtual Reality

Qian Zhou Autodesk Research Toronto, Canada qian.zhou@autodesk.com George Fitzmaurice Autodesk Research Toronto, Canada george.fitzmaurice@autodesk.com Fraser Anderson Autodesk Research Toronto, Canada fraser.anderson@autodesk.com

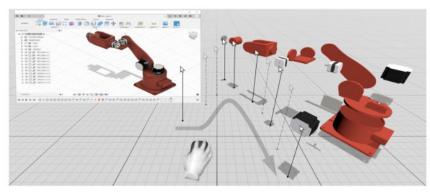


Figure 1: We investigate a technique that integrates a desktop mouse into VR to support productive knowledge work. Our approach uses Depth-Adaptive Cursor, a 2D-mouse driven pointing technique for 3D selection with depth-adaptation that continuously interpolates the cursor depth by inferring what users intend to select based on the cursor position, the viewpoint, and the selectable objects. Vertically dropped lines and arrow are added for illustration of depth.

ABSTRACT

Virtual Reality (VR) has potential for productive knowledge work, however, midair pointing with controllers or hand gestures does not offer the precision and comfort of traditional 2D mice. Directly integrating mice into VR is difficult as selecting targets in a 3D space is negatively impacted by binocular rivalry, perspective mismatch, and improperly calibrated control-display (CD) gain. To address these issues, we developed Depth-Adaptive Cursor. a

that Depth-Adaptive Cursor significantly improved performance compared with an existing mouse-based pointing technique without depth-adaption in terms of time (21.2%), error (48.3%), perceived workload, and user satisfaction.

CCS CONCEPTS

Human-centered computing → Pointing; Virtual reality.

Example: HCI (Interaction Technique)

Introduction

RW

System

User Study & Results

Discussion

Limitations & Future Work

Conclusion

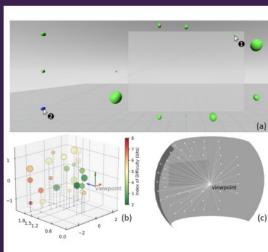
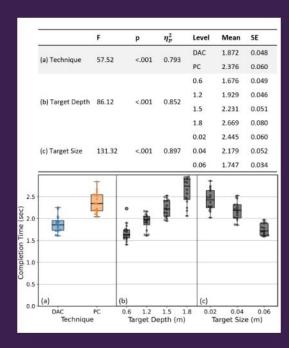


Figure 6: Setup of the 3D Pointing Task: (a) The task is to select an origin 2D target (a square on a 2D plane) and then again in a destination 3D target (a blue sphere in a 3D space). We use the combination of 2D and 3D targets to represent

Illustrate User Task



Statistical Results

Example: Al

Introduction

RW

Method

Implementation

Evaluation

Limitations & Discussion

Conclusion

Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

Taesung Park* Jun-Yan Zhu* Phillip Isola Alexei A. Efros Berkeley AI Research (BAIR) laboratory, UC Berkeley

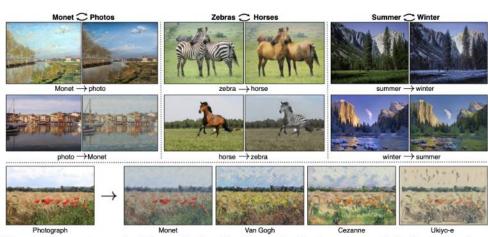


Figure 1: Given any two unordered image collections X and Y, our algorithm learns to automatically "translate" an image from one into the other and vice versa; (left) Monet paintings and landscape photos from Flickr; (center) zebras and horses from ImageNet; (right) summer and winter Yosemite photos from Flickr. Example application (bottom): using a collection of paintings of famous artists, our method learns to render natural photographs into the respective styles.

Example: Al

Introduction

RW

Method

Implementation

Evaluation

Limitations & Discussion

Conclusion

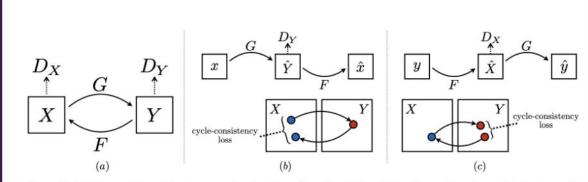


Figure 3: (a) Our model contains two mapping functions $G: X \to Y$ and $F: Y \to X$, and associated adversarial discriminators D_Y and D_X . D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_X and F. To further regularize the mappings, we introduce two cycle consistency losses that capture the intuition that if

$$\begin{split} \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) &= \mathbb{E}_{y \sim p_{\text{data}}(y)}[\log D_Y(y)] \\ &+ \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))], \end{split} \tag{1}$$

$$\begin{split} \mathcal{L}_{\text{cyc}}(G, F) &= \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] \\ &+ \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]. \end{split}$$

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F),$$

Example: Al

Introduction

RW

Method

Implementation

Evaluation

Limitations & Discussion

Conclusion

Adapted from Stanford CS197

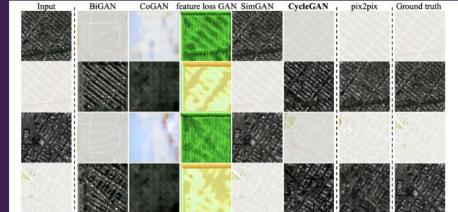


Figure 6: Different methods for mapping aerial photos↔maps on Google Maps. From left to right: input, BiGAN/ALI [7, 9], CoGAN [32], feature loss + GAN, SimGAN [46], CycleGAN (ours), pix2pix [22] trained on paired data, and ground truth.

| Loss | Map → Photo % Turkers labeled <i>real</i> | Photo → Map % Turkers labeled <i>real</i> |
|--------------------|--|--|
| CoGAN [32] | $0.6\% \pm 0.5\%$ | $0.9\% \pm 0.5\%$ |
| BiGAN/ALI [9, 7] | $2.1\% \pm 1.0\%$ | $1.9\% \pm 0.9\%$ |
| SimGAN [46] | $0.7\% \pm 0.5\%$ | $2.6\% \pm 1.1\%$ |
| Feature loss + GAN | $1.2\% \pm 0.6\%$ | $0.3\% \pm 0.2\%$ |
| CycleGAN (ours) | $26.8\% \pm 2.8\%$ | $23.2\% \pm 3.4\%$ |

Table 1: AMT "real vs fake" test on maps \leftrightarrow aerial photos at 256×256 resolution.

| Loss | Per-pixel acc. | Per-class acc. | Class IOU |
|--------------------|----------------|----------------|-----------|
| CoGAN [32] | 0.40 | 0.10 | 0.06 |
| BiGAN/ALI [9, 7] | 0.19 | 0.06 | 0.02 |
| SimGAN [46] | 0.20 | 0.10 | 0.04 |
| Feature loss + GAN | 0.06 | 0.04 | 0.01 |
| CycleGAN (ours) | 0.52 | 0.17 | 0.11 |
| pix2pix [22] | 0.71 | 0.25 | 0.18 |

Table 2: FCN-scores for different methods, evaluated on Cityscapes labels→photo.

| Loss | Per-pixel acc. | Per-class acc. | Class IOU |
|--------------------|----------------|----------------|-----------|
| CoGAN [32] | 0.45 | 0.11 | 0.08 |
| BiGAN/ALI [9, 7] | 0.41 | 0.13 | 0.07 |
| SimGAN [46] | 0.47 | 0.11 | 0.07 |
| Feature loss + GAN | 0.50 | 0.10 | 0.06 |
| CycleGAN (ours) | 0.58 | 0.22 | 0.16 |

| Loss | Per-pixel acc. | Per-class acc. | Class IOU |
|----------------------|----------------|----------------|-----------|
| Cycle alone | 0.22 | 0.07 | 0.02 |
| GAN alone | 0.51 | 0.11 | 0.08 |
| GAN + forward cycle | 0.55 | 0.18 | 0.12 |
| GAN + backward cycle | 0.39 | 0.14 | 0.06 |
| CycleGAN (ours) | 0.52 | 0.17 | 0.11 |

Table 4: Ablation study: FCN-scores for different variants of our method, evaluated on Cityscapes labels—photo.

| Loss | Per-pixel acc. | Per-class acc. | Class IOU | |
|----------------------|----------------|----------------|-----------|--|
| Cycle alone | 0.10 | 0.05 | 0.02 | |
| GAN alone | 0.53 | 0.11 | 0.07 | |
| GAN + forward cycle | 0.49 | 0.11 | 0.07 | |
| GAN + backward cycle | 0.01 | 0.06 | 0.01 | |
| CycleGAN (ours) | 0.58 | 0.22 | 0.16 | |

Table 5: Ablation study: classification performance of photo→labels for different losses, evaluated on Cityscapes.

method, on the other hand, can produce translations that are often of similar quality to the fully supervised pix2pix.

Table 1 reports performance regarding the AMT perceptual realism task. Here, we see that our method can fool participants on around a quarter of trials, in both the

Typical gold standard: conference

Computer Science, unlike other fields, is a conference-oriented field.

There are a small set of top-tier conferences for each area. These are generally known to be the venues that publish the best work in the area.

There also exist a variety of second-tier and other conferences, which are less prestigious and often easier to get into.

Journals, and conference-journal hybrids, fit into this category too.

Work-in-progress venues

You can only publish a research result once. Conferences and journals are known as **archival**, meaning that they are archived permanently in the academic record.

There also exist a variety of **non-archival venues** that are intended for feedback and exposure.

Workshops

Posters

Demos

arXiv.org

Work-in-progress venues

You can only publish a research result once. Conferences and journals are known as **archival**, meaning that they are archived permanently in the academic record.

There also exist a variety of **non-archival venues** that are intended for feedback and exposure.

Workshops

Posters

Demos

arXiv.org

Top conferences and journals

- Databases and Data Management:
 - VLDB, SIGMOD, ICDE
- Data Mining:
 - o KDD, WWW
- → HCI:
 - CHI, UIST, InfoVis
- Machine Learning:
 - NeurlPS (formerly NIPS), ICML, ICLR, AAAI
 - CVPR (vision), ACL (NLP), EMNLP (NLP)

Conference Ranks



Conference Data

| | Searc | h: eccv | eccv | |
|--|-------------|---------|-----------|--|
| Name | ↓ Abbrv. ↓1 | Rank 11 | Source 11 | |
| European Conference on Computer Vision | ECCV | Α | ERA | |
| European Conference on Computer Vision | ECCV | A1 | Qualis | |
| Name | Abbrv. | Rank | Source | |

Research metrics

- Research metrics are quantitative measurements designed to evaluate research outputs and their impacts.
 - journal, a journal article, a book, book chapter or the overall research productivity
- Metrics include a variety of measures and statistical methods for assessing the quality and broader impact of scientific and scholarly research.

Research metrics

They fall into 2 categories, bibliometrics and altmetrics.

- **Bibliometrics** are the traditional **citation** based metrics. They are based on citation counts, counting how many times a publication has been cited in another publication.
- Altmetrics are web based metrics. They are used for measuring the attention or the interest of a scholarly work on various types of online platforms that includes social media, research blogs, forums, etc.

The use of both types of metrics contribute to evaluate and understand the impact of a research

Access & Citations

Example:

https://www.nature

.com/articles/s422

56-019-0048-x

67k 1990 2329

Article Accesses Web of Science CrossRef

Online attention



This article is in the 99th percentile (ranked 1,007th) of the 355,676 tracked articles of a similar age in all journals and the 96th percentile (ranked 2nd) of the 27 tracked articles of a similar age in *Nature Machine Intelligence*

View more on Altmetric

Metrics for Journals: Journal Impact Factor

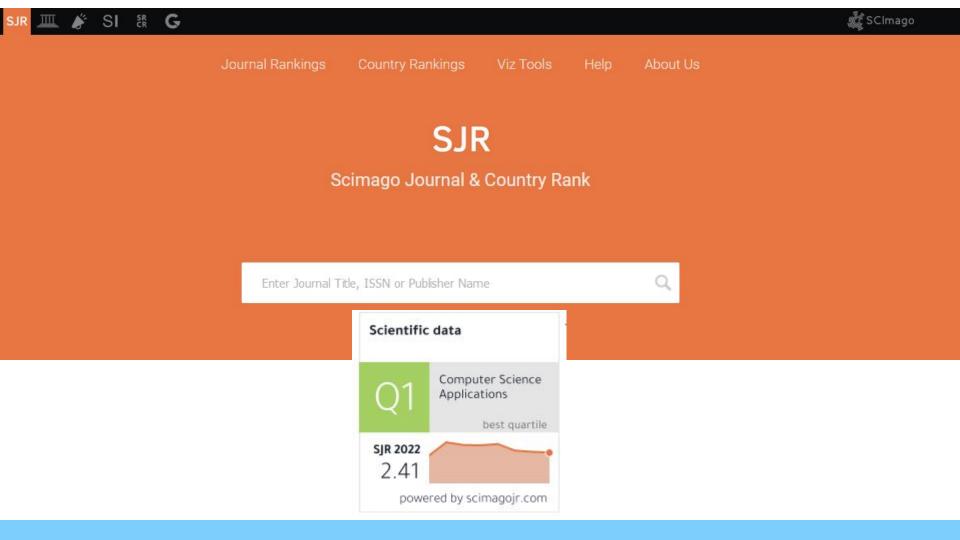
The most established metric for journal is the Journal Impact factor or JIF. It applies to academic journals. It's the very first metric designed for journal evaluation.

The main element of the JIF calculation are the average number of citations a paper receives- in a journal-in a given period of time. When a journal has an impact factor of 5 it means that in the last three years, this journal averaged 5 citations per published article.

- This metric is also described an indicator of a journal citedness.
- Journals are arranged in a ranking order from high to low JIF in their relevant field categories.
- JIF is calculated from the selected journals indexed in a source, for the JIF, its source is the major database Web of Science that contains the science, social sciences and art and humanities indexes.
- The ranking of journals are presented in quartiles and percentiles. Some journals appear in more than one categories.

Other Journal Metrics

- The <u>Scimago Journal Rank or SJR</u> that differentiates a **citation coming from an important journal into its formula**.
- The immediacy index that indicates journals that publish cutting-edge research.
- The Eigenfactor score that removes self-citations and calculates article citations in a journal in the past five years.
- The CiteScore that includes citations to all documents, not only articles and reviews, published in a journal in the past 3 years



What research is not

- I. Figure out what to do.
- 2. Do it.
- 3. Publish.

What research is

Research is an iterative process of exploration, not a linear path from idea to result [Gowers 2000]

Thesis vs. Research Paper

Research Paper

Limited space

Specialized context

Tackles a specific question

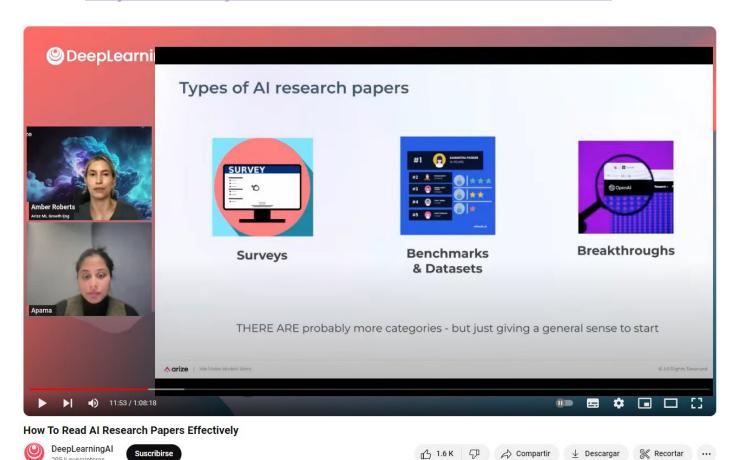
Thesis

"Unlimited" space

Broad context

Tackles a large question

https://www.youtube.com/watch?v=K6Wui3mn-ul





Tipos de publicaciones y métricas

Prof. Rosa Paccotacya Yanque

rypaccotacya@ucsp.edu.pe

E. P. de Ciencia de la Computación - UCSP

How to do a literature review

- 1. Pick your favorite academic search engine (e.g., Google scholar) and start with keywords
- 2. Find 3-5 recent and highly cited papers
 - From reputable venues and by reputable institution/author If you find a survey paper, start from the survey paper
- 3. Do the first pass to identify key papers and researchers that these works cite
- 4. Track down these papers/researchers
- Iterate as needed