AI VIETNAM
All-in-One Course

# How to Write a Research Paper 3

*(Introduction and Related Work Sections)*

Vinh Dinh Nguyen
PhD in Computer Science

*Year 2023*

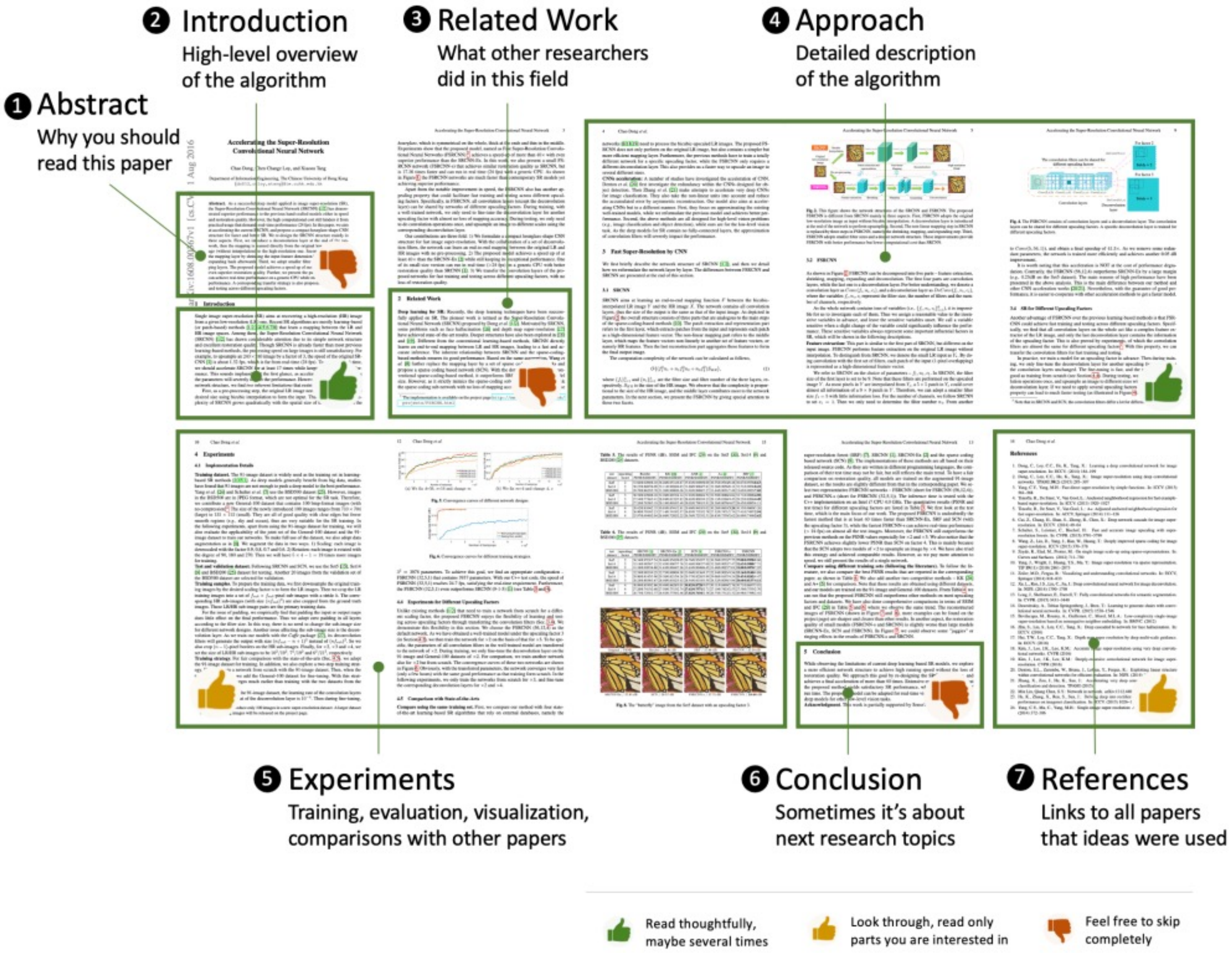
# Outline

- How to Write Related Work Section
- How to Write Introduction Section
- Assignment

# Outline

- How to Write Related Work Section
- How to Write Introduction Section
- Assignment

# How Do Researcher Spend Their Time


1 Abstract
Why you should read this paper
2 Introduction
High-level overview of the algorithm
3 Related Work
What other researchers did in this field
4 Approach
Detailed description of the algorithm
5 Experiments
Training, evaluation, visualization, comparisons with other papers
6 Conclusion
Sometimes it's about next research topics
7 References
Links to all papers that ideas were used
Read thoughtfully, maybe several times
Look through, read only parts you are interested in
Feel free to skip completely

# Schedule

**Week 1**
How to write the Proposed Method section

↓

**Week 2**
**How to write the Experimental Results section**

↓

**Week 3**
How to write the Introduction and Related Work sections

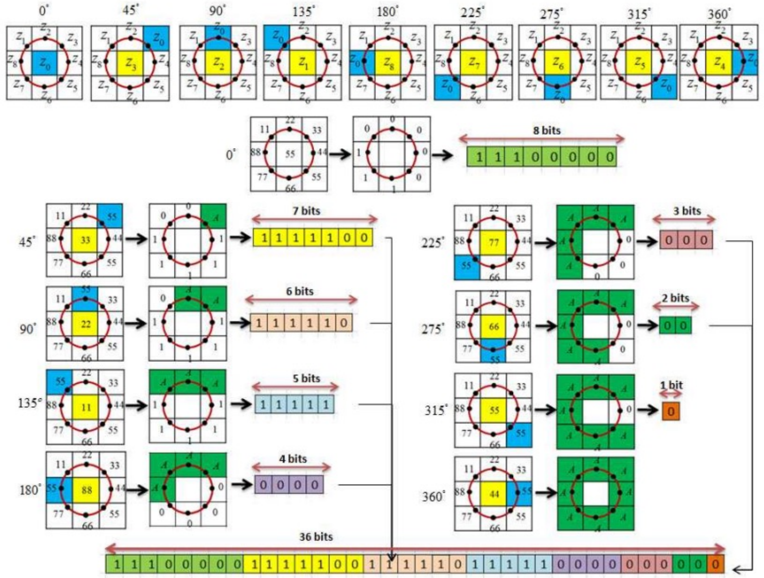
↓

**Week 4**
How to write the Conclusion, Abstract and Title sections

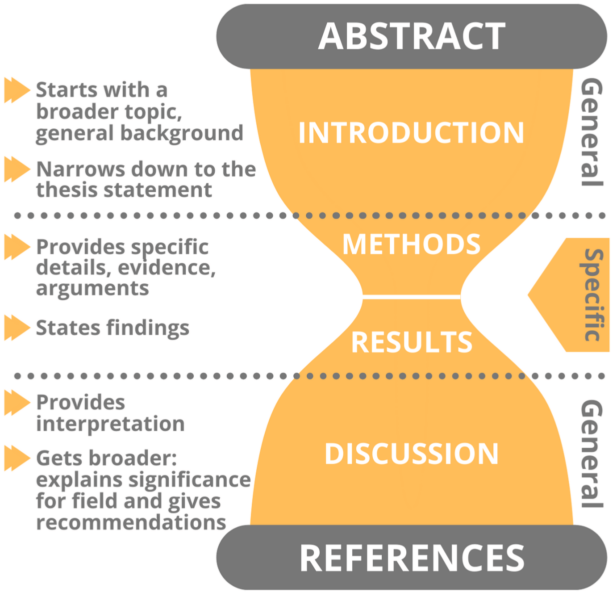
Trick and Tip for Answering the Reviewers Comments or Remarks

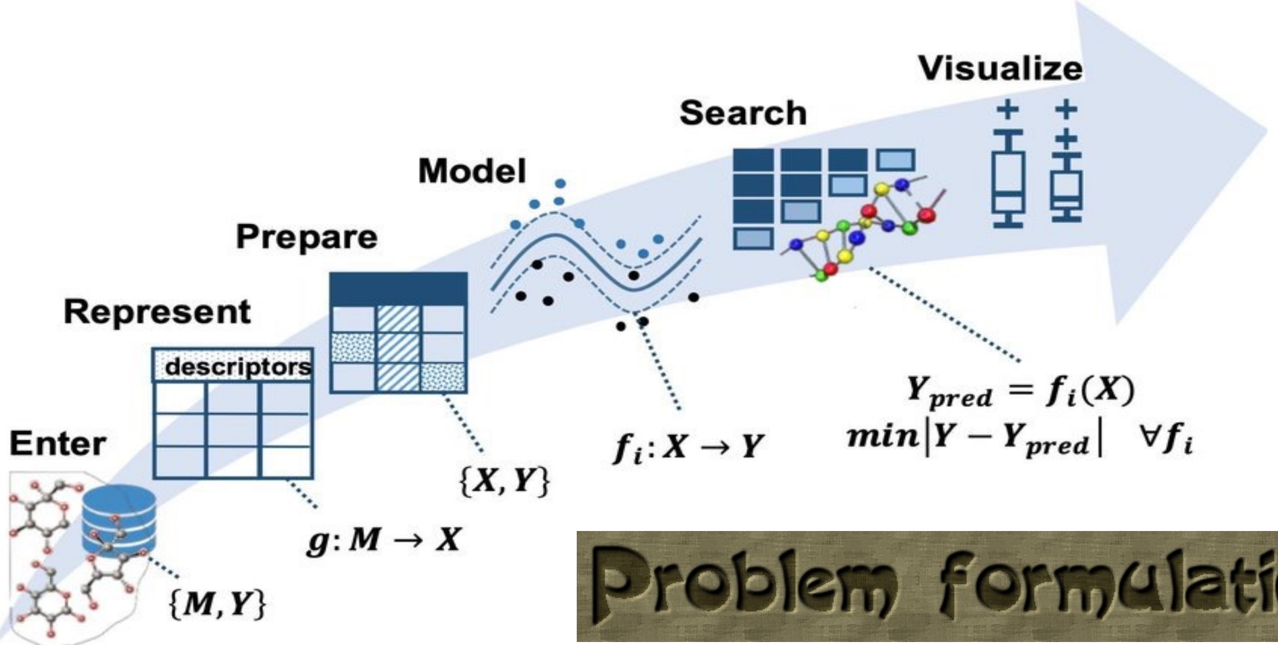
How to write the Reponse Letter

# How to Write Proposed Method




## RESEARCH PAPER STRUCTURE




$$
\begin{cases}
F^{(1)} = -C_{\mathrm{de}}G^{(1)}_{\mathrm{de}} + C_l G^{(1)}_l + C_r G^{(1)}_r + C_{\mathrm{to}}G^{(1)}_{\mathrm{to}} + C_b G^{(1)}_b \\
\qquad + C_d G^{(1)}_d + C_{\mathrm{rd}}G^{(1)}_{\mathrm{rd}} + C_m G^{(1)}_m + C_{\mathrm{pos}}G^{(1)}_{\mathrm{pos}} \\
\ldots \\
F^{(k)} = -C_{\mathrm{de}}G^{(k)}_{\mathrm{de}} + C_l G^{(k)}_l + C_r G^{(k)}_r + C_{\mathrm{to}}G^{(k)}_{\mathrm{to}} + C_b G^{(k)}_b \\
\qquad + C_d G^{(k)}_d + C_{\mathrm{rd}}G^{(k)}_{\mathrm{rd}} + C_m G^{(k)}_m + C_{\mathrm{pos}}G^{(k)}_{\mathrm{pos}} \\
\ldots \\
F^{(N)} = -C_{\mathrm{de}}G^{(N)}_{\mathrm{de}} + C_l G^{(N)}_l + C_r G^{(N)}_r + C_{\mathrm{to}}G^{(N)}_{\mathrm{to}} + C_b G^{(N)}_b \\
\qquad + C_d G^{(N)}_d + C_{\mathrm{rd}}G^{(N)}_{\mathrm{rd}} + C_m G^{(N)}_m + C_{\mathrm{pos}}G^{(N)}_{\mathrm{pos}} .
\end{cases}
$$

Problem formulation

Solution formulation

# STEP 1
## Sơ đồ thuật toán

Left Image → Cost Computation
Right Image → Cost Computation
Cost Computation → Winner Take All → Disparity Map

# CASE STUDY: STEREO MATCHING

# STEP 2
## Mô hình hoá công thức

Left Image → Cost Computation
Right Image → Cost Computation
Cost Computation → Winner Take All → Disparity Map

$$e(p,d) = \min(|I_l(x,y) - I_r(x-d,y)|, \sigma)$$

$$E(p,d) = \frac{\sum_{q \in N(p)} w(p,q) e(q,d)}{\sum_{q \in N(p)} w(p,q)}$$

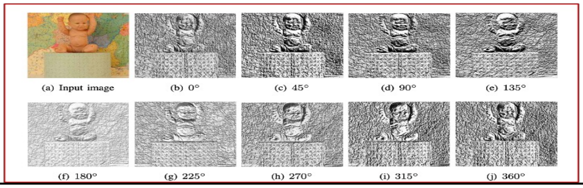
$$d(p) = \underset{d \in [0,\cdots,D-1]}{\arg\min} \; E(p,d),$$

$$\mathrm{NCC}(x,y,d) = \frac{\sum_{(x,y)\in w} I_l(x,y) \cdot I_r(x-d,y)}{\sqrt{\sum_{(x,y)\in w} I_l^2(x,y) \cdot \sum_{(x,y)\in w} I_r^2(x-d,y)}}.$$

$$d_p = \underset{d \in D}{\arg\min}\, C'(p,d).$$

# STEP 3
## Trực quan hoá ý tưởng

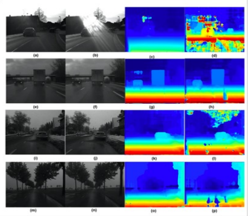
Left Image → Cost Computation
Right Image → Cost Computation
Cost Computation → Winner Take All → Disparity Map



(a) Input image (b) 0° (c) 45° (d) 90° (e) 135°
(f) 180° (g) 225° (h) 270° (i) 315° (j) 360°

# STEP 4
## Đánh giá kết quả ngắn gọn

Left Image → Cost Computation
Right Image → Cost Computation
Cost Computation → Winner Take All → Disparity Map

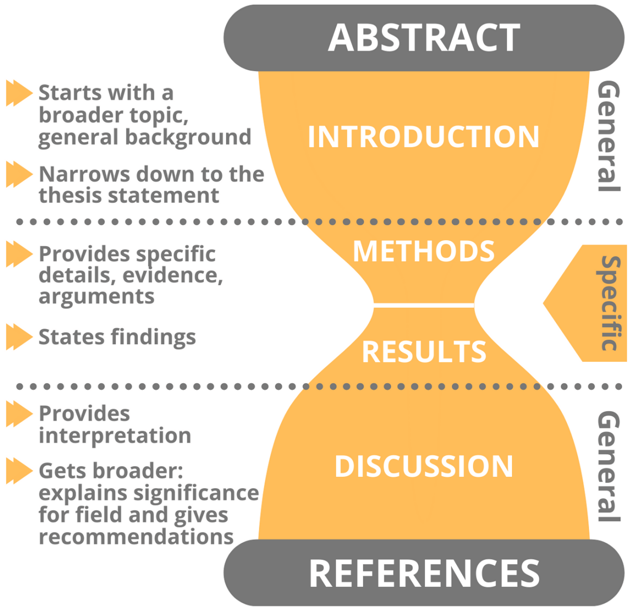


**Cầu nối cho phần "Experimental Results and Dicussion"**

AI VIETNAM
All-in-One Course

# Experimental Results

## RESEARCH PAPER STRUCTURE



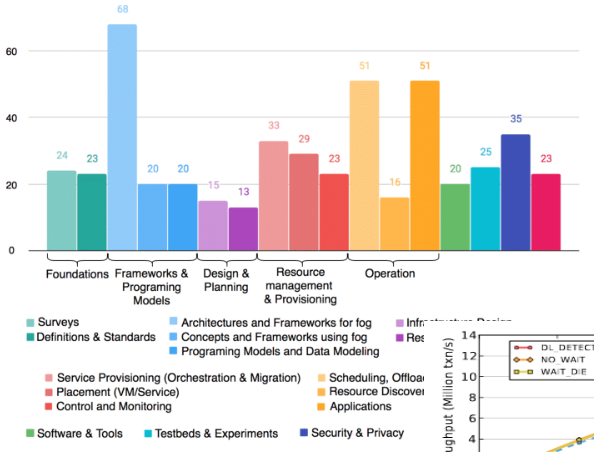




TABLE II
BENCHMARK OF STATE-OF-THE-ART AND PROPOSED METHODS (DR AND MDD)

| Author | DR % | Detection direction | Processing time(ms) | MDD (m) |
|---|---|---|---|---|
| Nedevschi [3] | 94 | Front and rear | 100 | 90 |
| Yi-Min Tsai [10] | 97.1 | Rear | 970 | 140 |
| Toulminet [17] | 90 | Rear | 95 | 100 |
| Yaqian Li [20] | 91 | Front and rear | 50 | 105 |
| Xu [24] | 85 | Front and rear | 155.5 | 70.58 |
| Nedevschi [34] | 96.69 | Front and rear | 23 | N/A |
| S. Sivaraman [47] | 93.75 | Rear | 57 | 88 |
| R. Wang [48] | 95.6 | Rear | 510 | 89 |
| Y.M. Chan [49] | 85 | Front and rear | 47 | 105 |
| Bin-dai [50] | 93 | Rear | 96.9 | 90 |
| B. Leibe [51] | 98.5 | Rear | 2150.5 | 110 |
| **Proposed method** | **96.08** | **Front and rear** | **48** | **140** |

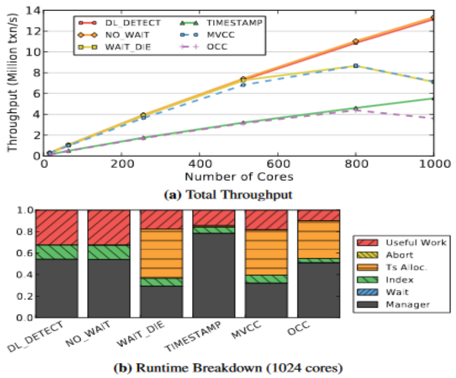
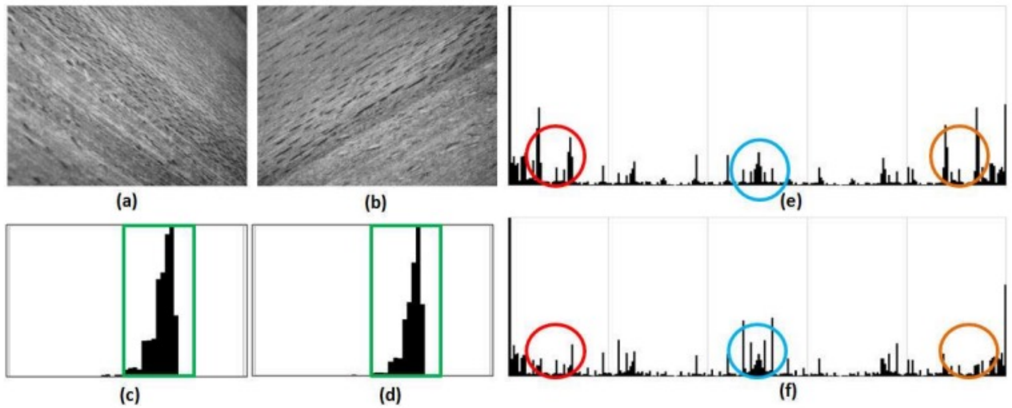
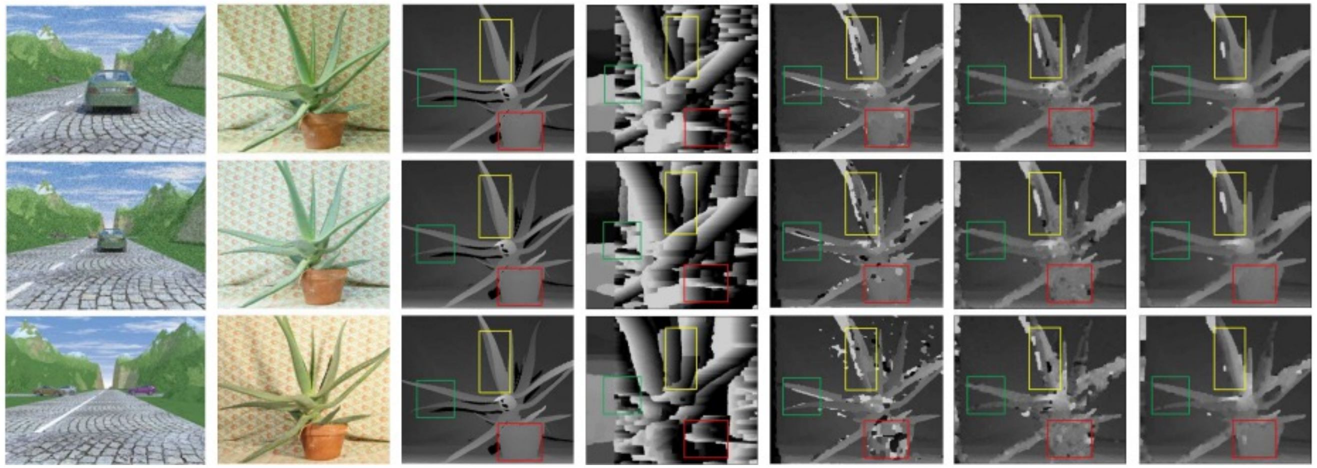



**Figure 8: Read-only Workload** – Results for a read-only YCSB workload.



# Experimental Results

## STEP 1: Research Questions

- **Research Questions:**
  - RQ1: How effective is the proposed method in improving the performance of the local stereo matching under difficult conditions?
  - RQ2: How effective is the proposed method in comparison to the existing method such as SAD, NCC, AD?
  - RQ3: How does the performance of the proposed method by varying the Cost Size, Aggregation Cost Size?

## STEP 2: Dataset, Existing Method, and Evaluation Metrics

### Evaluation Metrics

Root mean square (RMS) [24] and percentage of bad pixel matching are often used to estimate the difference between a stereo matching result and the ground truth disparity at time index $t$

$$\mathrm{RMS}(t) = \left( \frac{1}{P} \sum_{(x,y)} |d_t(x,y) - g_t(x,y)|^2 \right)^{\frac{1}{2}} \quad (9)$$

where $P$ is the total number of pixels in the image, and $d_t(x,y)$ and $g_t(x,y)$ are the intensity values at pixel $(x,y)$ of the computed disparity image and the ground truth disparity image, respectively. In addition, the average root mean square (ARMS) is used to evaluate the error of stereo matching results for long sequences

$$\mathrm{ARMS}_N = \frac{1}{N} \left( \frac{1}{P} \sum_{(x,y)} |d_t(x,y) - g_t(x,y)|^2 \right)^{\frac{1}{2}} \quad (10)$$

where $N$ is the total number of images in the testing sequences.

The percentage of bad pixel matching in the matching result and the ground truth disparity at time index $t$ are often defined as follows:

$$B(t) = \frac{1}{P} \sum_{(x,y)} \left( |d_t(x,y) - g_t(x,y)| > \alpha \right). \quad (11)$$

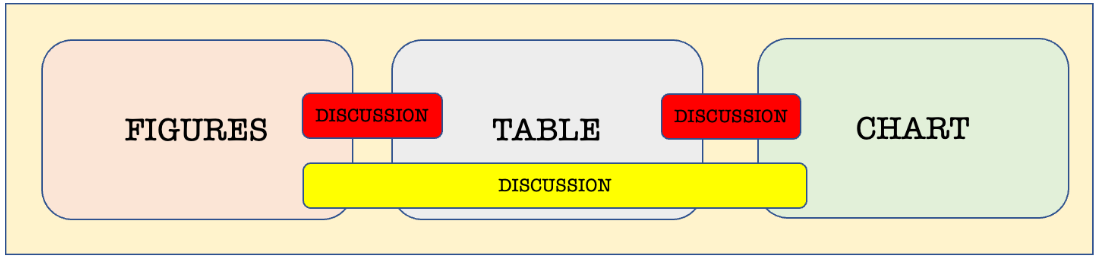
ps://ieeexplor

### KITTI Dataset

| | Method | Setting | Code | D1-bg | D1-fg | D1-all |
|---|---|---|---|---|---|---|
| 1 | MoCha-Stereo | | | 1.36 % | 2.43 % | 1.53 % |
| 2 | DiffuVolume | | | 1.35 % | 2.51 % | 1.54 % |
| 3 | GANet+ADL | | | 1.38 % | 2.38 % | 1.55 % |
| 4 | Selective-IGEV | | | 1.33 % | 2.61 % | 1.55 % |

## STEP 3: Answers for each Research Question

RQ1: How effective is the proposed method in improving the performance of the local stereo matching under difficult conditions?




## Discuss Limitations of The Proposed Method

**Accuracy**

**Processing Time**

**Complexity**

# How to write Related Works

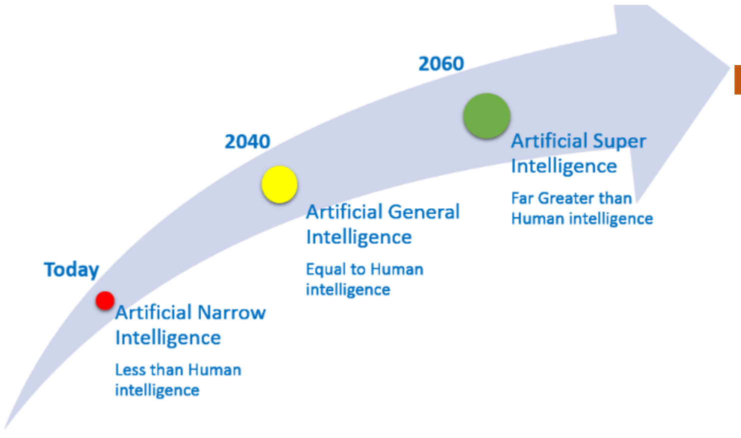



*Figure 4: Future evolution of Artificial Intelligence*

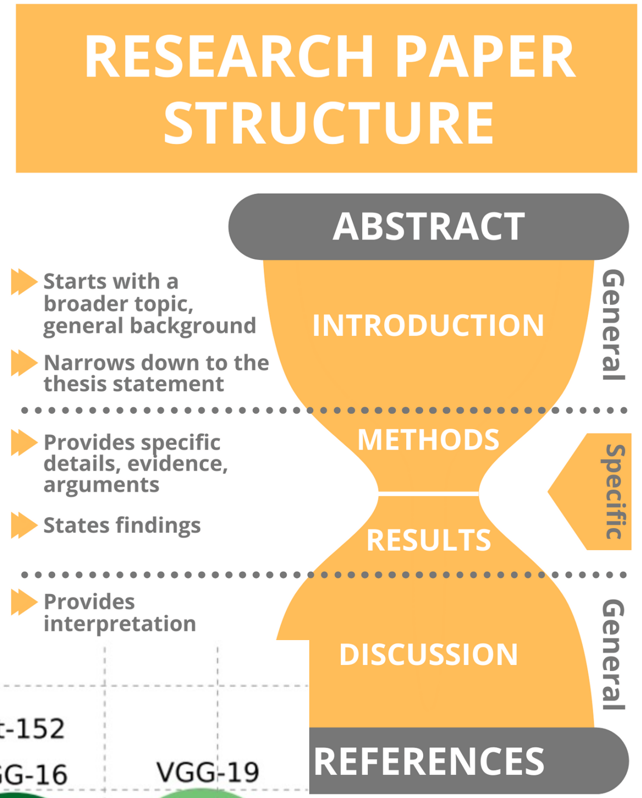
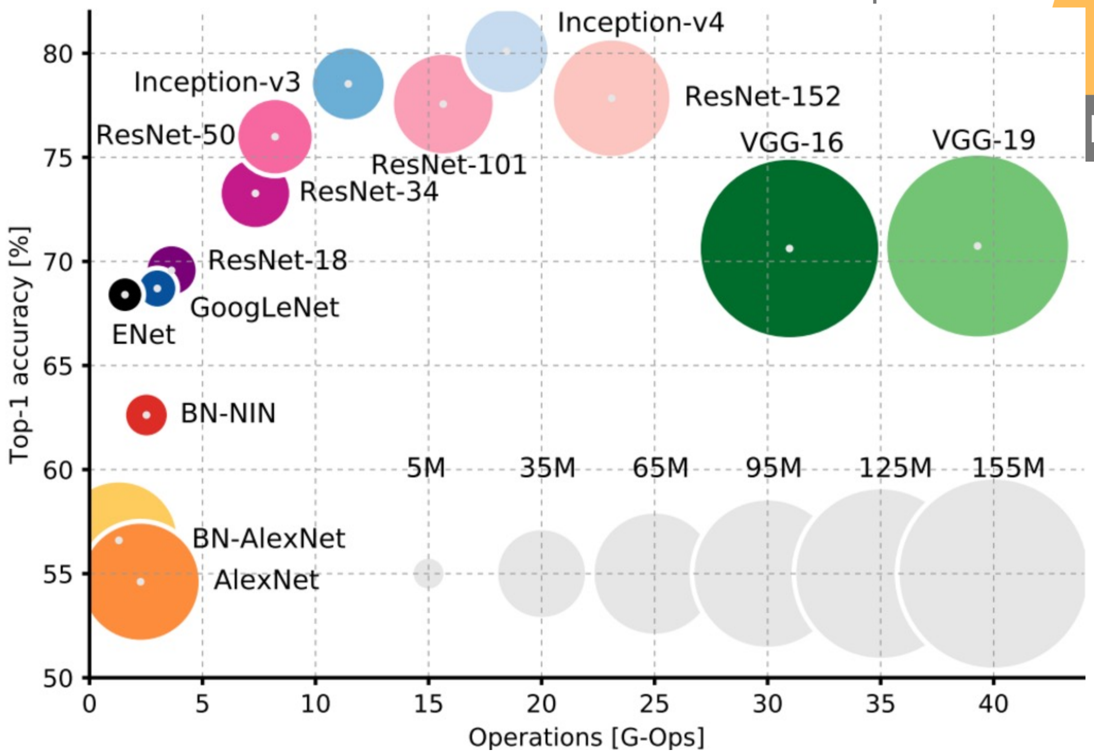



TABLE I
COMPARISON OF STEREO-VISION-BASED VDDE SYSTEMS

| System | Stereo configuration | MDD | Stereo matching | Feature/Technique | Other |
|---|---|---|---|---|---|
| Nedevschi (2004) [3] | $N/A$ | $90m$ | SAD | 3–D points grouping | Four additional sensors are used. Processing time: 100ms/frame |
| Pauplin (2005) [6] | $N/A$ | $N/A$ | $N/A$ | Fly algorithm, Parisian approach | 3-D point is randomly generated in ROI. Two CCD cameras are used. Processing time: 300ms/frame |
| Ruicheck (2004) [7] | $N/A$ | $N/A$ | Genetic-based | Genetic algorithm | Multi-level searching and edge information.Processing time: 400ms/frame |
| Knoeppel (2000) [15] | Focal length: $12mm$ | $150m$ | Cross correlation | 3–D point extracting and clustering; Kalman filter | CMOS cameras |
| Bensrhair (2002) [16] | Baseline: $95cm$ | $N/A$ | DP applied on edge points | 3–D edge shapes of roads, symmetry | CCD cameras used to obtain disparity; Processing time: 960ms/frame |
| Toulminet (2006) [17] | Focal length: $6mm$ | $100m$ | DP on edge points | 3–D vertical features; symmetry; image correlation | Two stereo cameras used with monocular pattern analysis. Processing time: 95ms/frame |
| Jung (2007) [18] | Baseline: $30cm$ | $60m$ | Edge feature | Lane recognition; edge feature correlation | CMOS cameras |
| Huh (2008) [19] | Baseline: $15cm$ | $65m$ | Corner feature | Corner features; epipolar constraint; Kalman filter | Developed for highway circumstance |
| Southall (2009) [21] | Baseline: 7 inches ($17.78cm$) | $100m$ | SAD | Edge features; Kalman filter | NTSC cameras |
| Kormann (2010) [22] | $N/A$ | $30m$ | SAD | Mean–shift clustering of plane fitted segments; U/V disparity | 3–D data from disparity map is used to detect vehicles as cuboid. Processing time: 390ms/frame. |
| Hwang (2009) [23] | Baseline: $150m$ | $N/A$ | Edge feature based matching | Corner features; cross correlation; Kalman filter | Developed for highway circumstance. Processing time: 10ms/frame |
| Xu (2009) [24] | Baseline: $96cm$ | $N/A$ | SAD | Image symmetrical move | Symmetrical matching. Processing time: 155.5ms/frame. Detection rate: 85% |
| Chiu (2010) [25] | Baseline: $20cm$ | $50m$ | Horizontal and vertical line segments | Edge features; line segments | CMOS cameras; Asynchronous binocular system. Processing time: 30ms/frame. Detection rate: 90% |
| Franke (2011) [28] | Focal length: 830 pixels, Baseline: 35cm | $N/A$ | SGM | Segmentation of the Dynamic Stixel World. | A Stixel is build by fusion of stereo and motion information. Processing time: 25ms/frame |



# What You Need to Know

*Skills need to know*

*How to find Research Papers*

↓

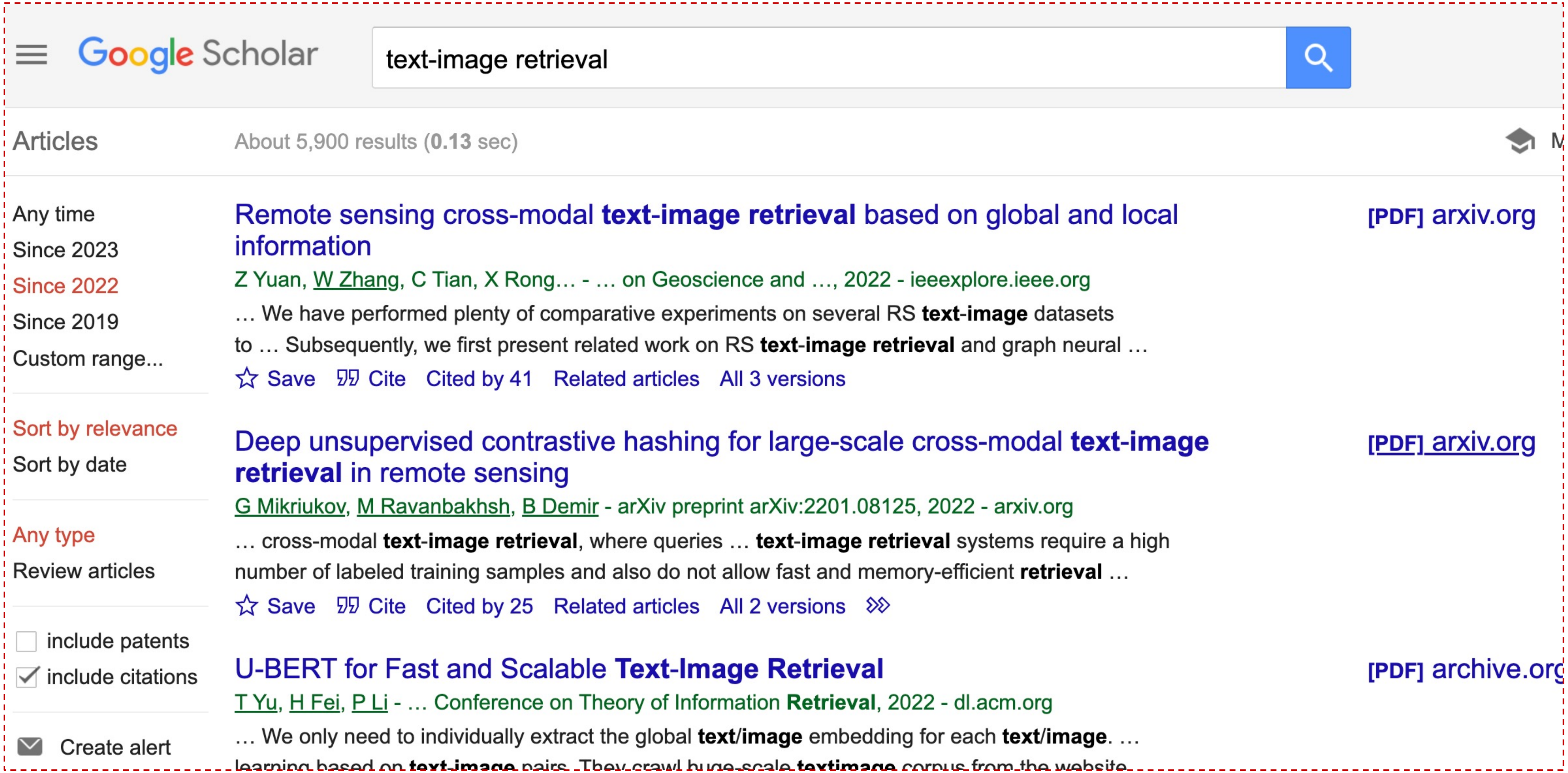
*How to read Research Papers*

↓

*How to Summary Research Papers*

AI VIETNAM
All-in-One Course

# How to find a research paper



# How to find a research paper

## What to look ?

- Specific
- Year
- Journal
- Author

## Where to look ?

IEEE Xplore® Digital Library

Google

Springer

Google Scholar

AUTHENTIC

PubMed

ScienceDirect

AI VIETNAM
All-in-One Course

# How to Read a Research Paper

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 24, NO. 2, FEBRUARY 2014 263

## Step 1:

Read Paper Title and answer the following questions?

1. What is the research topic?
2. Is it closed to your research interest?
3. Is it a new research approach?
4. General idea of what the research is about
5. Does the title use keywords that are relevant to the topic?
6. Does the title catch your attention or pique your interest in the research?

## Support Local Pattern and Its Application to Disparity Improvement and Texture Classification

Vinh Dinh Nguyen, Dung Duc Nguyen, Thuy Tuong Nguyen, Vinh Quang Dinh, and Jae Wook Jeon, *Member, IEEE*

***Abstract*—The local binary pattern (LBP) and its variants have been widely investigated in many image processing and computer vision applications due to their robust ability to capture local image structures and their computational simplicity. The existing LBPs extract local structure information by establishing a relationship between the central pixel and its adjacent pixels. However, most LBPs miss the relationship among all of the pixels in the local region. Therefore, this paper proposes a novel model to establish this relationship by introducing a support LBP. The proposed model improves the performance of the existing LBP methods and results in lower sensitivity to illumination changes and radiometric variations. Moreover, the proposed model has been successfully investigated in two applications: disparity map generation and texture classification. For disparity map generation, the proposed model reduces the root mean square (RMS) error by 23.6% (in Baby1 dataset, Middlebury), and 16.58% (in Aloe dataset, Middlebury) as compared with the standard LBP under radiometric variation conditions. Moreover, the proposed model reduces the RMS by 28.11% as compared with the standard LBP under the Gaussian noise condition in the ESATS dataset. For texture classification applications, the proposed model improves the classification results from 96.26% to 98.13% on the Outext database, from 88.03% to 91.41% on the Xu database, and from 94.00% to 96.67% on the KTH-TIPS database as compared with the completed LBP.**

method that extracts the structure of the local region based on the differences in intensity between the central pixel and its adjacent pixels. Generally, we can classify the LBP and its variants into five main categories: multiscale analysis, handling rotation, handling color, complementary descriptors, and feature selection and learning, as in [3]. Texture classification has been one of the most popular and successful applications that has used LBP during the past few years.

Texture classification is an application that is used to assign an unknown texture to the set of predefined textures. The assignment process can be based on the training of VZMR8, VZ_Joint [4], [5], or free training [6]. VZMR8 uses a predefined texton library that is constructed from a training set and a set of filter banks in order to classify the unknown texture. A texture dataset is divided into two sets, one of which is for a sample model and one that is used for testing during the free training. In conventional LBP [1], a uniform pattern is used to extract the texture information. It achieves very good accuracy in texture classification by combining multiple-scale LBPs, which work well in the case of straight edges or low curvature edges. However, this method is not successful in cases with

Decision

AI VIETNAM
All-in-One Course

# How to Read a Research Paper

## Step 2:

Look at the authors of the paper and answer the following questions?

1. Who are the authors of the paper?
2. Are their main research topics?
3. Do they have any related publications?
4. Is there a corresponding author indicated, and what are their contact details?

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 24, NO. 2, FEBRUARY 2014 263

### Support Local Pattern and Its Application to Disparity Improvement and Texture Classification

Vinh Dinh Nguyen, Dung Duc Nguyen, Thuy Tuong Nguyen, Vinh Quang Dinh, and Jae Wook Jeon, *Member, IEEE*

***Abstract*—The local binary pattern (LBP) and its variants have been widely investigated in many image processing and computer vision applications due to their robust ability to capture local image structures and their computational simplicity. The existing LBPs extract local structure information by establishing a relationship between the central pixel and its adjacent pixels. However, most LBPs miss the relationship among all of the pixels in the local region. Therefore, this paper proposes a novel model to establish this relationship by introducing a support LBP. The proposed model improves the performance of the existing LBP methods and results in lower sensitivity to illumination changes and radiometric variations. Moreover, the proposed model has been successfully investigated in two applications: disparity map generation and texture classification. For disparity map generation, the proposed model reduces the root mean square (RMS) error by 23.6% (in Baby1 dataset, Middlebury), and 16.58% (in Aloe dataset, Middlebury) as compared with the standard LBP under radiometric variation conditions. Moreover, the proposed model reduces the RMS by 28.11% as compared with the standard LBP under the Gaussian noise condition in the ESATS dataset. For texture classification applications, the proposed model improves the classification results from 96.26% to 98.13% on the Outext database, from 88.03% to 91.41% on the Xu database, and from 94.00% to 96.67% on the KTH-TIPS database as compared with the completed LBP.**

method that extracts the structure of the local region based on the differences in intensity between the central pixel and its adjacent pixels. Generally, we can classify the LBP and its variants into five main categories: multiscale analysis, handling rotation, handling color, complementary descriptors, and feature selection and learning, as in [3]. Texture classification has been one of the most popular and successful applications that has used LBP during the past few years.

Texture classification is an application that is used to assign an unknown texture to the set of predefined textures. The assignment process can be based on the training of VZMR8, VZ_Joint [4], [5], or free training [6]. VZMR8 uses a predefined texton library that is constructed from a training set and a set of filter banks in order to classify the unknown texture. A texture dataset is divided into two sets, one of which is for a sample model and one that is used for testing during the free training. In conventional LBP [1], a uniform pattern is used to extract the texture information. It achieves very good accuracy in texture classification by combining multiple-scale LBPs, which work well in the case of straight edges or low curvature edges. However, this method is not successful in cases with

Decision

# How to Read a Research Paper

## Step 2:

Look at the authors of the paper and answer the following questions?

1. Who are the authors of the paper?
2. Are their main research topics?
3. Do they have any related publications?
4. Is there a corresponding author indicated, and what are their contact details?

[37] C. D. Pantilie and S. Nedevschi, "Real time obstacle detection in complex scenarios using dense stereo vision and optical flow," in *Proc. 13th Int. IEEE Intell. Transp. Syst.*, 2010, pp. 439–444.
[38] M. Nishigaki and Y. Aloimonos, "Moving obstacle detection using cameras for driver assitance system," in *Proc. IEEE Intell. Veh. Symp.*, 2010, pp. 805–812.
[39] U. Franke, C. Rabe, H. Badino, and S. Gehrig, "6D-vision: Fustion of stereo and motion for robust environment perception," in *Proc. DAGM Conf. Pattern Recog.*, 2005, pp. 216–223.
[40] B. Jhne, *Digital Image Processing*, 6th ed. New York, NY, USA: Springer-Verlag, 2005, pp. 91–142.
[41] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.
[42] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA, USA: Addison-Wesley, 1989.
[43] R. Labayrade, D. Aubert, and D. J. P. Tarel, "Real time obstacle detection on non flat road geometry through v-disparity representation," in *Proc. IEEE Symp. Intell. Veh.*, 2002, vol. 2, pp. 646–651.
[44] C. Pocol, S. Nedevschi, and M. M. Meinecke, "Obstacle detection based on dense stereovision for urban ACC systems," in *Proc. Int. Workshop on Intell. Transp.*, 2008, pp. 13–18.
[45] E. K. P. Chong and S. H. Zak, *An Introduction to Optimization*, 2nd ed. Hoboken, NJ, USA: Wiley, 2001, pp. 187–212.
[46] T. Blickle and L. Thiele, "A mathematical analysis of tournament selection," in *Proc. Int. Conf. Genetic Algorithms*, 1995, pp. 9–16.
[47] S. Sivaraman and M. M. Trivedi, "Active learning based robust monocular vehicle detection for on-road safety systems," in *Proc. IEEE Symp. Intell. Veh.*, 2009, pp. 399–404.
[48] C.-C. R. Wang and J.-J. J. Lien, "Automatic vehicle detection using local features—A statistical approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 83–96, Mar. 2008.
[49] Y.-M. Chan, S.-S. Huang, L.-C. Fu, and P.-Y. Hsiao, "Vehicle detection under various lighting conditions by incorporating particle filter," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, 2007, pp. 534–539.
[50] B. Dai, "A vehicle detection method via symmetry in multi-scale windows," in *Proc. IEEE Conf. Ind. Electron. Appl.*, 2007, pp. 1827–1831.
[51] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *Int. J. Comput. Vis.*, vol. 77, no. 1–3, pp. 259–289, May 2008.
[52] S. Bauer, S. Kohler, K. Doll, and U. Brunsmarn, "FPGA-GPU architecture for kernel SVM pedestrian detection," in *Proc. IEEE Int. Comput. Vis. Pattern Recog. Workshops*, 2010, pp. 61–68.
[53] T. Machida and T. Naito, "GPU & CPU cooperative accelerated pedestrian and vehicle detection," in *Proc. IEEE Int. Comput. Vis. Workshops*, 2011, pp. 506–513.

**Vinh Dinh Nguyen** received the B.S. degree (*magna cum laude*) in computer science from Nong Lam University, Ho Chi Minh City, Vietnam, in 2007 and the M.S. degree in electrical and computer engineering from Sungkyunkwan University, Suwon, Korea, in 2012. He is currently working toward the Ph.D. degree with the School of Information and Communication Engineering, Sungkyunkwan University.

His research interest include computer vision, image processing, and graphics processing unit computing.

AI VIETNAM
All-in-One Course

# How to Read a Research Paper

## Step 2:

Look at the authors of the paper and answer the following questions?

1. Who are the authors of the paper?
2. Are their main research topics?
3. Do they have any related publications?
4. Is there a corresponding author indicated, and what are their contact details?


IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 24, NO. 2, FEBRUARY 2014 263

Support Local Pattern and Its Application to Disparity Improvement and Texture Classification

Vinh Dinh Nguyen, Dung Duc Nguyen, Thuy Tuong Nguyen, Vinh Quang Dinh, and

Vinh Dinh Nguyen

> Home > Persons

[–] 2020 – today

2023

[j11] Vinh Dinh Nguyen, Thong Duc Trinh, Hoang Ngoc Tran: **A Robust Triangular Sigmoid Pattern-Based Obstacle Detection Algorithm in Resource-Limited Devices.** IEEE Trans. Intell. Transp. Syst. 24(6): 5936-5945 (2023)

[c11] Vinh Dinh Nguyen, Thanh Hoang Tran, Doan Thai Dang, Narayan C. Debnath: **Robust Vehicle Detection by Using Deep Learning Feature and Support Vector Machine.** AICV 2023: 149-157

[c10] Vinh Dinh Nguyen, Ngoc Phuong Ngo, Narayan C. Debnath: **Leaf Disease Detection in Blueberry Using Efficient Semi-supervised Learning Approach.** AICV 2023: 188-196

[c9] Bach Hoang Ngo, Dat Thanh Nguyen, Nhat-Tuong Do-Tran, Phuc Pham Huy Thien, Minh-Hung An, Tuan-Ngoc Nguyen, Loi Nguyen Hoang, Vinh Dinh Nguyen, Quang-Vinh Dinh: **Comprehensive Visual Features and Pseudo Labeling for Robust Natural Language-based Vehicle Retrieval.** CVPR Workshops 2023: 5409-5418

2022

[j10] Vinh Quang Dinh, Phuc Hong Nguyen, Vinh Dinh Nguyen: **Feature Engineering and Deep Learning for Stereo Matching Under Adverse Driving Conditions.** IEEE Trans. Intell. Transp. Syst. 23(7): 7855-7865 (2022)

[j9] Hyung-Joon Jeon, Vinh Dinh Nguyen, Tin Trung Duong, Jae Wook Jeon: **A Deep Learning Framework for Robust and Real-Time Taillight Detection Under Various Road Conditions.** IEEE Trans. Intell. Transp. Syst. 23(11): 20061-20072 (2022)

Decision

# How to Read a Research Paper

## Step 3:

Read the abstract of the paper and answer the following questions?

1. What methods or approaches were used in the research?
2. What are the most important findings or results of the study?
3. Which techniques that you known/or unknown?
4. What conclusions or implications are drawn from the findings?
5. What is the overall purpose or goal of the research?

*Abstract*—The local binary pattern (LBP) and its variants have been widely investigated in many image processing and computer vision applications due to their robust ability to capture local image structures and their computational simplicity. The existing LBPs extract local structure information by establishing a relationship between the central pixel and its adjacent pixels. However, most LBPs miss the relationship among all of the pixels in the local region. Therefore, this paper proposes a novel model to establish this relationship by introducing a support LBP. The proposed model improves the performance of the existing LBP methods and results in lower sensitivity to illumination changes and radiometric variations. Moreover, the proposed model has been successfully investigated in two applications: disparity map generation and texture classification. For disparity map generation, the proposed model reduces the root mean square (RMS) error by 23.6% (in Baby1 dataset, Middlebury), and 16.58% (in Aloe dataset, Middlebury) as compared with the standard LBP under radiometric variation conditions. Moreover, the proposed model reduces the RMS by 28.11% as compared with the standard LBP under the Gaussian noise condition in the ESATS dataset. For texture classification applications, the proposed model improves the classification results from 96.26% to 98.13% on the Outext database, from 88.03% to 91.41% on the Xu database, and from 94.00% to 96.67% on the KTH-TIPS database as compared with the completed LBP.

Decision

# How to Read a Research Paper

## Step 4:

Read the conclusion of the paper and answer the following questions?

1. What are the main findings and results presented in the conclusion?
2. How do the findings answer the research question or address the research objectives?
3. Are there any unexpected or surprising results discussed in the conclusion?
4. Do the conclusions align with the research hypothesis or initial goals?
5. Are there any limitations or weaknesses of the study mentioned in the conclusion?
6. What are the practical implications of the findings?
7. Are there any recommendations or suggestions for future research provided?

### V. Conclusion

This paper analyzed LBP from the viewpoint of the local structure in order to investigate the feasibility and effectiveness of the support binary pattern for various existing binary patterns and applications. The support binary pattern successfully establishes the relationship of a pixel from various directions in the local region. Robust experiments on two common applications, disparity map generation and texture classification, demonstrate that the proposed model improves the performance of current local pattern methods. The main contributions of this paper include the following.

1) Proposal of a new model to establish the relationships among all of the pixels in the local region (support local pattern).
2) Successful application of the proposed model to disparity map generation and texture classification to improve their performances in the case of radiometric variations and Gaussian noise.

Decision

# How to Read a Research Paper

## Step 4:

Read the conclusion of the paper and answer the following questions?

1. What are the main findings and results presented in the conclusion?
2. How do the findings answer the research question or address the research objectives?
3. Are there any unexpected or surprising results discussed in the conclusion?
4. Do the conclusions align with the research hypothesis or initial goals?
5. Are there any limitations or weaknesses of the study mentioned in the conclusion?
6. What are the practical implications of the findings?
7. Are there any recommendations or suggestions for future research provided?

Conclusion

3) Easy application of the proposed model to other local patterns to improve performance, while preserving the simplicity of the original methods.

Support local pattern is a general model. Therefore, it should be slightly redesigned to satisfy the main requirements of a specific application, such as a rotation invariant of texture classification or robustness to noise for disparity map generation. The effect of the proposed model for other local patterns on various applications, including face detection, face recognition, and human detection, will be investigated in the near future.

Decision

# How to Read a Research Paper

## Step 5:

Read the introduction of the paper and answer the following questions?

1. What is the overall purpose or goal of the research?
2. What is the background or context of the research?
3. What are previous related research that has been done?
4. What are the citation of each previous work (extremely important for you)
5. Is there a review of related work or literature in the introduction?
6. Are there any notable trends or debates in the field mentioned?
7. What is the theoretical framework or approach used in the research?
8. What is the scope and focus of the research?
9. Does the introduction engage your interest and make you want to continue reading the paper?

***Abstract*—The local binary pattern (LBP) and its variants have been widely investigated in many image processing and computer vision applications due to their robust ability to capture local image structures and their computational simplicity. The existing LBPs extract local structure information by establishing a relationship between the central pixel and its adjacent pixels. However, most LBPs miss the relationship among all of the pixels in the local region. Therefore, this paper proposes a novel model to establish this relationship by introducing a support LBP. The proposed model improves the performance of the existing LBP methods and results in lower sensitivity to illumination changes and radiometric variations. Moreover, the proposed model has been successfully investigated in two applications: disparity map generation and texture classification. For disparity map generation, the proposed model reduces the root mean square (RMS) error by 23.6% (in Baby1 dataset, Middlebury), and 16.58% (in Aloe dataset, Middlebury) as compared with the standard LBP under radiometric variation conditions. Moreover, the proposed model reduces the RMS by 28.11% as compared with the standard LBP under the Gaussian noise condition in the ESATS dataset. For texture classification applications, the proposed model improves the classification results from 96.26% to 98.13% on the Outext database, from 88.03% to 91.41% on the Xu database, and from 94.00% to 96.67% on the KTH-TIPS database as compared with the completed LBP.**

***Index Terms*—Completed local binary pattern (CLBP), local binary pattern (LBP), local derivative pattern (LDP), support local pattern.**

### I. Introduction

In recent years, the local binary pattern (LBP) [1] has been investigated in many applications, such as texture classification, face detection, face recognition, medical image analysis, gender classification, human detection, and object detection [2], [3], due to its robustness to illumination changes and its simplicity of computation. LBP is a nonparametric method that extracts the structure of the local region based on the differences in intensity between the central pixel and its adjacent pixels. Generally, we can classify the LBP and its variants into five main categories: multiscale analysis, handling rotation, handling color, complementary descriptors, and feature selection and learning, as in [3]. Texture classification has been one of the most popular and successful applications that has used LBP during the past few years.

Texture classification is an application that is used to assign an unknown texture to the set of predefined textures. The assignment process can be based on the training of VZMR8, VZ_Joint [4], [5], or free training [6]. VZMR8 uses a predefined texton library that is constructed from a training set and a set of filter banks in order to classify the unknown texture. A texture dataset is divided into two sets, one of which is for a sample model and one that is used for testing during the free training. In conventional LBP [1], a uniform pattern is used to extract the texture information. It achieves very good accuracy in texture classification by combining multiple-scale LBPs, which work well in the case of straight edges or low curvature edges. However, this method is not successful in cases with complicated shapes such as crossing boundaries or corners, as shown in [7]. The dominant LBP [7] is proposed in order to solve this issue by considering the most frequently occurring patterns in texture images. An analysis in [8] showed that the LBP does not retain global spatial information, so a new LBP with global matching is proposed using the LBP variance. To find another component to supplement LBP, Zhang *et al*. [9] introduced a monogenic-LBP by combining LBP and monogenic signal theory. Recently, Guo *et al*. [6] demonstrated that LBP with a sign operator captures more discriminative information than LBP with a magnitude operator. A completed local binary pattern (CLBP) is proposed based on these operators and the center pixel value. Motivated by the order derivative of the LBP, the local derivative pattern (LDP) [10] is proposed to capture more detailed information by introducing a high-order derivative. However, if the order is greater than three, LDP is more sensitive to noise than LBP. The local ternary pattern (LTP) [11] is proposed in order to decrease noise dependence, especially near the uniform region. LTP provides three values for encoding [−1, 0, 1], with the help of a predefined threshold, instead of the two-value encoding [0, 1] used in LBP.

The greatest advantage of LBP is its robustness with respect to illumination change. Many stereo methods have been

Manuscript received September 24, 2012; revised January 02, 2013; accepted February 08, 2013. Date of publication March 27, 2013; date of current version February 4, 2014. This work was supported in part by the Ministry of Knowledge Economy, Korea, under Information Technology Research Center NIPA-2012-(H0301-12-3001) and by PRCP through the National Research Foundation of Korea, funded by MEST under Grant 2012-0005861. This paper was recommended by Associate Editor A. Vetro.

V. D. Nguyen, D. D. Nguyen, V. Q. Dinh, and J. W. Jeon are with the School of Information and Communication Engineering, Sungkyunkwan University, Suwon 440-746, Korea (e-mail: vinhnd@skku.edu; nddunga3@skku.edu; dqvinh@skku.edu; jwjeon@yurim.skku.ac.kr).

T. T. Nguyen is with Image Mining Group, Institut Pasteur Korea, Seongnam 463-400, Korea (e-mail: ntthuy@ip-korea.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2013.2254898

Decision

# How to Read a Research Paper

## Step 6:

## Read the Experimental Results section of the paper and answer the following questions?

1. What are the key findings or results presented in the experimental results section?
2. Do the results address the research question or hypothesis posed in the introduction?
3. Are there any unexpected or surprising results discussed in the section?
4. Are there any statistical analyses or tests mentioned, and if so, what were the results?
5. Do the results support or contradict the prior literature discussed in the introduction?
6. Are there any figures, tables, or visuals, and do they enhance your understanding of the results?
7. Which existing methods they mentioned for comparison?
8. Which datasets they used for evaluation?
9. Which conditions they conducted for testing?

texture classification by abandoning the microstructure. Although the local structure information is discarded, LBC obtains the same performance as LBP in rotation-invariant cases. In addition, a completed local count binary (CLBC) is introduced based on LBC to compete with CLBP in [26]. LBC (or LBP) is sensitive to random noise and quantization noise

IV. Experimental Results

To assess the strength of the proposed method, we conducted robust experiments on two applications: disparity map generation and texture classification. The publicly available Middlebury [27] image dataset and the EISATS [28] synthetic vehicle sequences are used to evaluate the effectiveness of the



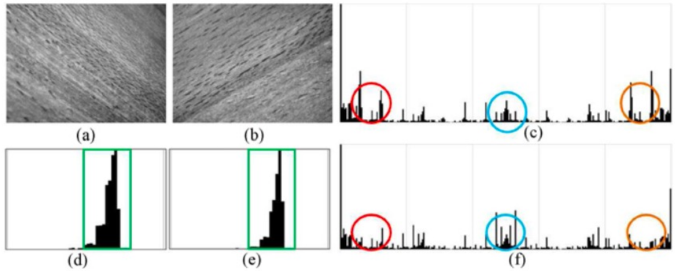


Fig. 9. (a) and (b) Two texture images. (c) and (d) SLBP_TC histograms of (a) and (b), respectively. (e) and (f) LBP histograms of (a) and (b), respectively.

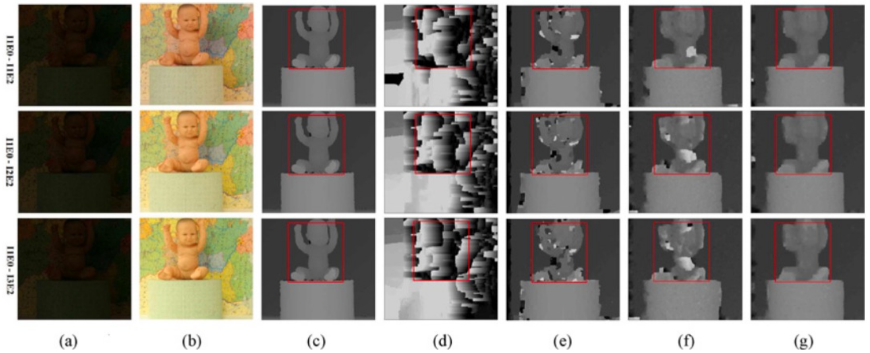


Fig. 10. Result of various stereo methods on Baby1 image pairs under various illuminations and exposures. (a) Left image. (b) Right image. (c) Gr truth. (d) AD. (e) ANCC. (f) LBP. (g) SLBP.

Decision

# How to Read a Research Paper

## Step 6:

Read the Experimental Results section of the paper and answer the following questions?

1. What are the key findings or results presented in the experimental results section?
2. Do the results address the research question or hypothesis posed in the introduction?
3. Are there any unexpected or surprising results discussed in the section?
4. Are there any statistical analyses or tests mentioned, and if so, what were the results?
5. Do the results support or contradict the prior literature discussed in the introduction?
6. Are there any figures, tables, or visuals, and do they enhance your understanding of the results?
7. Which existing methods they mentioned for comparision?
8. Which datasets they used for evaluation?
9. Which conditions they coducted for testing?

CLASSIFICATION RATE (%) ON TC10 AND TC12 USING DIFFERENT SCHEMES (OUTEXT DATASET)

| | (N, R)=(8,1) | | | | (N, R)=(16,2) | | | | (N, R)=(24,3) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TC10 | TC12 | | Average | TC10 | TC12 | | Average | TC10 | TC12 | | Average |
| | | *t* | *h* | | | *t* | *h* | | | *t* | *h* | |
| LTP | 76.06 | 62.56 | 63.42 | 67.34 | 96.11 | 85.2 | 85.87 | 89.06 | 98.64 | 92.59 | 91.52 | 94.25 |
| VAR(*P*, *R*) | 90 | 62.93 | 64.35 | 72.42 | 86.71 | 63.47 | 67.26 | 72.48 | 81.66 | 58.98 | 64.18 | 68.80 |
| LBP(*riu*2, *P*, *R*)/VAR(*P*, *R*) | 96.56 | 79.31 | 78.08 | 84.65 | 97.84 | 85.76 | 85.54 | 89.38 | 98.15 | 87.13 | 87.08 | 90.79 |
| M-LBP | 81.21 | 78.85 | 77.6 | 79.22 | 83.55 | 80.1 | 79.35 | 81.00 | 85.56 | 84.32 | 83.67 | 85.41 |
| CLBP | 96.56 | 90.3 | 92.29 | 93.05 | 98.72 | 93.54 | 93.91 | 95.39 | 98.93 | 95.32 | 94.53 | 96.26 |
| SCLBP | 95.45 | 89.45 | 92.88 | 92.59 | 98.05 | 92.77 | 94.21 | 95.01 | 98.85 | 95.25 | 94.45 | 96.18 |
| CLBC | 97.16 | 89.79 | 92.92 | 93.29 | 98.54 | 93.26 | 94.07 | 95.29 | 98.78 | 94.00 | 93.24 | 95.67 |
| SCLBP_TC (*t* = 0) | 97.25 | 89.75 | 92.93 | 93.31 | 98.45 | 93.28 | 93.98 | 95.24 | 99.05 | 93.95 | 93.27 | 95.42 |
| SCLBP_TC (*t* = 1) | 97.45 | 90.25 | 93.02 | 93.57 | 99.68 | 94.25 | 94.75 | 95.89 | 99.25 | 95.77 | 99.45 | 96.82 |
| **SCLBP_TC(*t* = 3)** | **98.25** | **92.45** | **94.25** | **94.98** | **99.36** | **95.05** | **96.47** | **96.96** | **99.45** | **96.68** | **98.25** | **98.13** |
| SCLBP_TC (*t* = 5) | 96.45 | 89.25 | 92.15 | 92.62 | 97.89 | 94.02 | 93.25 | 93.05 | 98.26 | 94.86 | 97.22 | 96.78 |

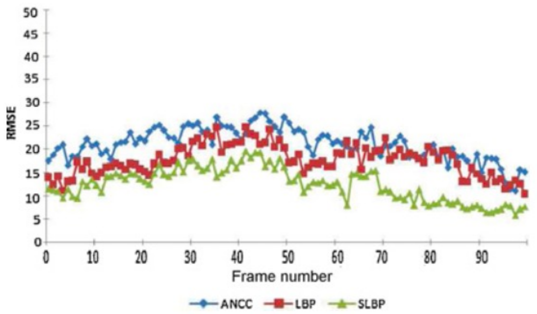
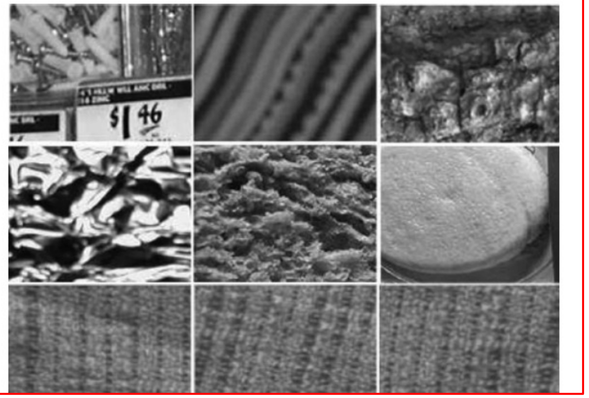



Fig. 14. RMS error of various stereo methods using the synthetic sequence with Gaussian noise.



Decision

# How to Read a Research Paper

## Step 6:

Read the Experimental Results section of the paper and answer the following question?

1. What are the key findings or results presented in the experimental results section?
2. Do the results address the research question or hypothesis posed in the introduction?
3. Are there any unexpected or surprising results discussed in the section?
4. Are there any statistical analyses or tests mentioned, and if so, what were the results?
5. Do the results support or contradict the prior literature discussed in the introduction?
6. Are there any figures, tables, or visuals, and do they enhance your understanding of the results?
7. Which existing methods they mentioned for comparision?
8. Which datasets they used for evaluation?
9. Which conditions they coducted for testing?

TABLE III

CLASSIFICATION RATE (%) ON XU DATASET USING DIFFERENT SCHEMES

| *T* | (*N*, *R*)=(8,1) | | | | | (*N*, *R*)=(16,2) | | | | | (*N*, *R*)=(24,3) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 20 | 15 | 10 | 5 | Average | 20 | 15 | 10 | 5 | Average | 20 | 15 | 10 | 5 | Average |
| LTP | 51.40 | 50.08 | 45.06 | 37.70 | 46.06 | 77.45 | 75.66 | 74.28 | 70.88 | 74.57 | 83.25 | 80.58 | 79.46 | 75.55 | 79.71 |
| M-LBP | 83.55 | 81.46 | 79.42 | 69.15 | 78.39 | 85.56 | 82.23 | 79.47 | 70.55 | 79.45 | 87.43 | 85.66 | 84.05 | 82.78 | 84.98 |
| CLBP | 86.80 | 82.40 | 78.93 | 71.20 | 78.83 | 87.00 | 85.02 | 80.53 | 73.60 | 81.54 | 90.25 | 88.59 | 87.34 | 85.94 | 88.03 |
| SCLBP | 86.25 | 82.21 | 77.68 | 70.05 | 79.05 | 86.55 | 84.33 | 79.21 | 72.11 | 80.55 | 88.75 | 87.56 | 86.59 | 85.02 | 86.98 |
| CLBC | 87.25 | 83.11 | 79.43 | 72.66 | 80.61 | 87.69 | 86.98 | 82.46 | 74.55 | 82.92 | 91.25 | 89.66 | 87.55 | 86.18 | 88.66 |
| SCLBP_TC ($t = 0$) | 87.28 | 83.45 | 79.55 | 73.00 | 80.82 | 87.25 | 87.05 | 82.59 | 75.04 | 82.98 | 92.02 | 89.05 | 88.25 | 85.59 | 88.73 |
| **SCLBP_TC($t = 1$)** | **88.35** | **83.88** | **81.79** | **73.03** | **81.76** | **87.99** | **86.24** | **82.68** | **74.47** | **83.1** | **93.97** | **92.25** | **91.48** | **87.94** | **91.41** |
| SCLBP_TC ($t = 3$) | 87.55 | 82.58 | 80.24 | 71.25 | 80.41 | 86.59 | 86.25 | 81.97 | 74.48 | 83.1 | 82.45 | 93.05 | 90.45 | 86.02 | 88.89 |
| SCLBP_TC ($t = 5$) | 86.16 | 82.09 | 79.98 | 70.06 | 79.92 | 85.67 | 84.86 | 80.61 | 73.39 | 81.23 | 92.55 | 89.78 | 89.25 | 85.24 | 89.21 |

TABLE IV

CLASSIFICATION RATE (%) ON KTH-TIPS DATASET USING DIFFERENT SCHEMES

| *T* | (*N*, *R*)=(8,1) | | | | | (*N*, *R*)=(16,2) | | | | | (*N*, *R*)=(24,3) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 30 | 20 | 10 | 5 | Average | 30 | 20 | 10 | 5 | Average | 30 | 20 | 10 | 5 | Average |
| LTP | 72.06 | 66.8 | 60.73 | 54.44 | 63.50 | 85.33 | 81.64 | 77.56 | 75.23 | 79.94 | 90.65 | 89.24 | 87.63 | 84.33 | 87.96 |
| M-LBP | 88.25 | 87.18 | 80.27 | 75.32 | 82.75 | 92.44 | 90.55 | 84.88 | 75.46 | 85.83 | 94.56 | 92.35 | 88.26 | 85.88 | 90.26 |
| CLBP | 91.91 | 91.59 | 83.97 | 77.46 | 86.23 | 96.32 | 92.41 | 84.85 | 78.45 | 88.01 | 98.35 | 97.64 | 91.55 | 88.48 | 94.00 |
| SCLBP | 90.84 | 89.98 | 81.57 | 77.02 | 84.85 | 96.01 | 91.23 | 83.33 | 77.98 | 87.14 | 96.55 | 95.68 | 89.34 | 87.11 | 92.17 |
| CLBC | 92.08 | 92.49 | 85.25 | 78.53 | 87.09 | 96.66 | 93.54 | 86.25 | 79.04 | 88.87 | 98.25 | 98.33 | 92.00 | 89.45 | 94.51 |
| SCLBP_TC ($t = 0$) | 91.86 | 92.55 | 85.22 | 78.09 | 86.93 | 96.02 | 92.98 | 87.03 | 78.69 | 88.68 | 97.89 | 98.25 | 92.25 | 88.95 | 94.34 |
| **SCLBP_TC($t = 1$)** | **94.05** | **92.88** | **86.11** | **79.03** | **88.02** | **97.55** | **93.88** | **87.06** | **79.52** | **89.50** | **99.25** | **99.24** | **96.28** | **91.89** | **96.67** |
| SCLBP_TC ($t = 3$) | 93.68 | 91.56 | 82.98 | 78.25 | 86.62 | 96.14 | 92.63 | 86.24 | 78.36 | 88.34 | 98.36 | 97.85 | 95.33 | 90.64 | 95.55 |
| SCLBP_TC ($t = 5$) | 91.22 | 90.56 | 83.66 | 77.96 | 85.85 | 95.68 | 92.09 | 85.46 | 75.59 | 87.21 | 97.25 | 98.05 | 94.64 | 90.29 | 95.06 |

Decision

# How to Read a Research Paper

## Step 7:

Read the Proposed Method section of the paper and answer the following questions?

1. What is the research problem or objective?
2. What is the overall approach or framework used in the method?
3. What data preprocessing or cleaning steps are involved?
4. What specific techniques or algorithms are applied?
5. Are there any assumptions made in the method?
6. How are the parameters and hyperparameters chosen or tuned?
7. How does the proposed method compare to existing methods?
8. How reproducible is the method?

II. LBP AND THE PROPOSED MODEL

In this section, we briefly review common local patterns such as the LBP and the CLBP and their disadvantages. To address the limitations of these patterns, the SLBP model is then introduced.

*A. LBP*

LBP [1] is a nonparametric method that captures the local structures of images by comparing each pixel with eight neighbor pixels ($N = 8$) in a $3 \times 3$ window ($R = 1$). Given a center pixel $z_c = (x, y)$ in the image, the LBP code is generated as

$$LBP_{N,R}(z_c) = \sum_{n=1}^{N} f(z_n, z_c)2^n$$
$$f(z_n, z_c) = \begin{cases} 1, & I(z_n) \geq I(z_c) \\ 0, & I(z_n) < I(z_c) \end{cases} \quad (1)$$

where $I(z_c)$ and $I(z_n)$ are the gray values of the central pixel and the $n$th neighbor pixel, respectively. $R$ is the radius of the neighborhood. Moreover, $N$ neighborhood pixels can be located outside of the image grids. Their gray values can be calculated using bilinear interpolation, and the coordinate of $z_n$ is determined by

$$(x_n, y_n) = \left[x_c + R\cos\left(\frac{2\pi n}{N}\right), y_c - R\sin\left(\frac{2\pi n}{N}\right)\right]. \quad (2)$$

Another type of LBP is the uniform pattern. A pattern is considered uniform if the number of bitwise transitions between

*C. General Model of SLBP*

Analyzing LBP from the view point of the local structures, we propose a supporting local pattern in order to establish a relationship among the neighborhood pixels, while preserving the simplicity of the original method. Given a local structure ($N = 8$, $R = 1$), the LBP code is a binary string that is established by measuring the magnitude value between a center pixel and its neighborhood pixels. An intuitive question has been raised as to whether a binary string could be established to every pixel in the local region. This means that every pixel might act as a center pixel in the local region.

Fig. 2(c) shows the results of the binary strings that are generated from every pixel in the local region from Fig. 2(a). Thus, the discriminative information can be obtained by considering each pixel in the local region as the center pixel. The structure of the local region (original microscopic pattern configuration) is not destroyed for two reasons: 1) Every pixel position is not moved in the local region; 2) From the view of each pixel position, a binary string is established by considering its intensity value relative to the intensity value of other neighborhoods. That is exactly what LBP does to establish a binary string.

We do not consider the rotation invariant in the general SLBP model, because SLBP should be adaptively applied to specific applications based on the main characteristic of the application. For example, the disparity map generation does not require a rotation invariant, while the rotation invariant is an important property for texture classification. A simple example has been conducted to demonstrate the previous

Decision

AI VIETNAM
All-in-One Course

# How to Read a Research Paper

## Step 7:

Read the Proposed Method section of the paper and answer the following questions?

1. What is the research problem or objective?
2. What is the overall approach or framework used in the method?
3. What data preprocessing or cleaning steps are involved?
4. What specific techniques or algorithms are applied?
5. Are there any assumptions made in the method?
6. How are the parameters and hyperparameters chosen or tuned?
7. How does the proposed method compare to existing methods?
8. How reproducible is the method?

NGUYEN et al.: SUPPORT LOCAL PATTERN AND ITS APPLICATION TO DISPARITY IMPROVEMENT AND TEXTURE CLASSIFICATION 267

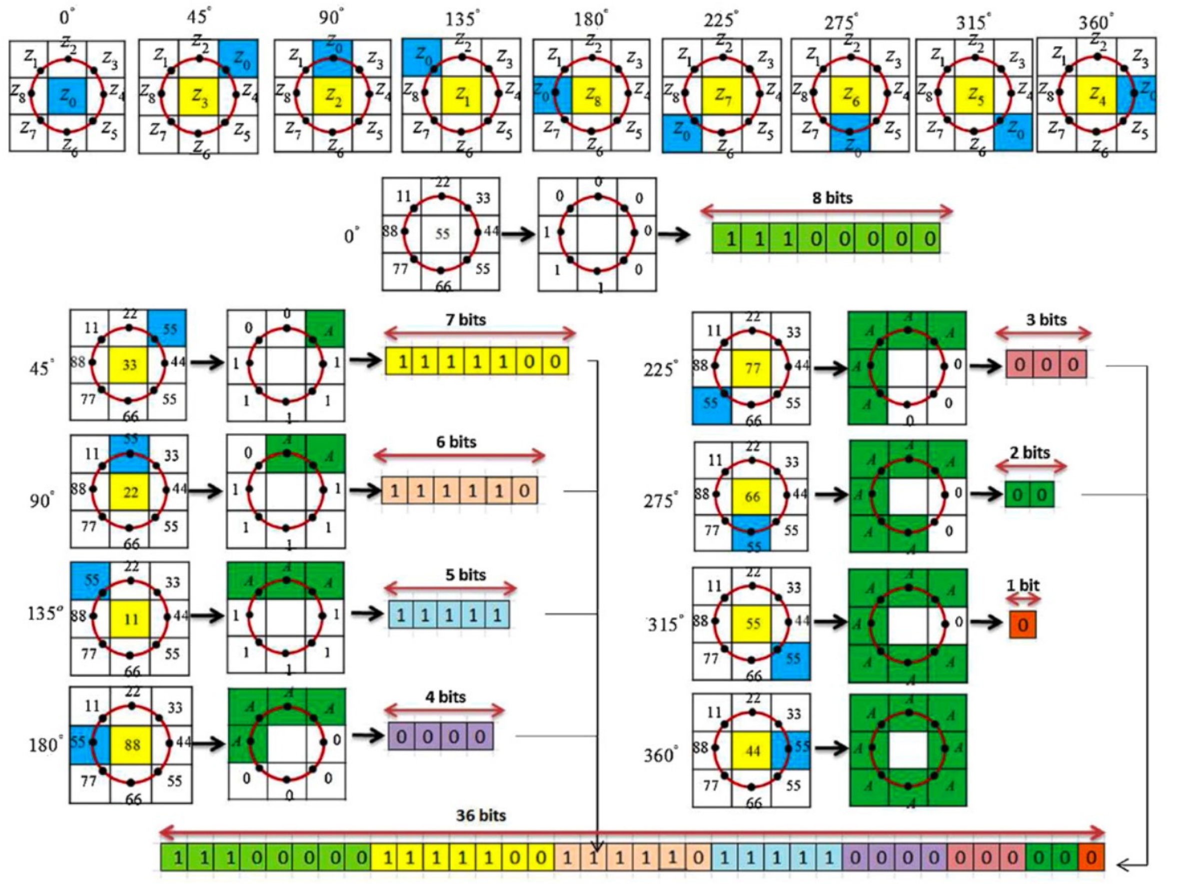



Fig. 4. Proposed data cost model (SLBP) for disparity map generation. *A* is denoted as a position of the pixel that has already established the relationship to the center pixel.

Decision

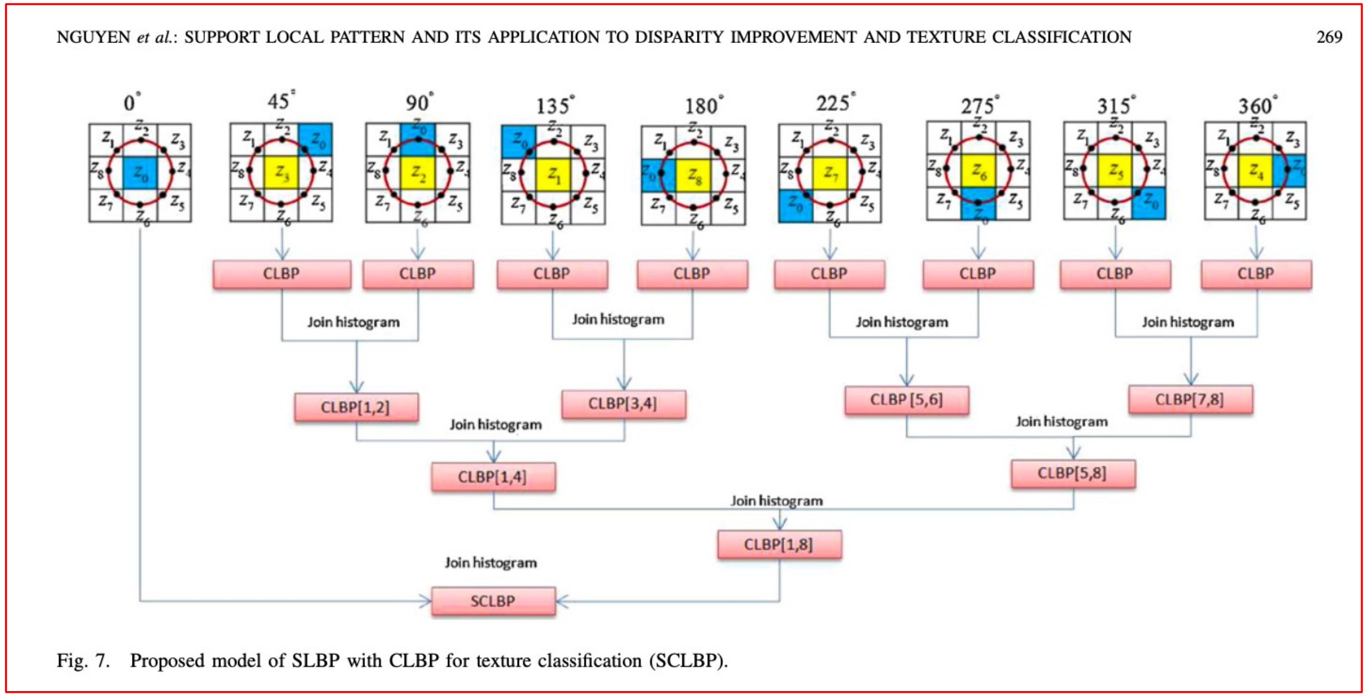
AI VIETNAM
All-in-One Course

# How to Read a Research Paper

## Step 7:

Read the Proposed Method section of the paper and answer the following questions?

1. What is the research problem or objective?
2. What is the overall approach or framework used in the method?
3. What data preprocessing or cleaning steps are involved?
4. What specific techniques or algorithms are applied?
5. Are there any assumptions made in the method?
6. How are the parameters and hyperparameters chosen or tuned?
7. How does the proposed method compare to existing methods?
8. How reproducible is the method?




Decision

AI VIETNAM
All-in-One Course

# How to Read a Research Paper

## Step 8:

Read the Related Work section of the paper and answer the following question?

1. What is the purpose of the "Related Work" section?
2. What are the key research topics or areas covered in the related work?
3. What are the most recent and relevant studies in the field?
4. How does the paper's research question or problem statement relate to the prior literature?
5. Are there any gaps or limitations in the existing research that the paper aims to address?
6. What are the main findings or insights from the related work?
7. How does the current research build upon or extend the existing literature?

TABLE I
COMPARISON OF VARIOUS STEREO METHODS FOR DISPARITY MAP GENERATION AND LBP METHODS FOR TEXTURE CLASSIFICATION AND FACE DETECTION

| Author | Algorithm | Descriptions | Type | Other |
|---|---|---|---|---|
| Texture classification | | | | |
| Guo *et al.* [6] | Completed LBP | Incorporation with different sign and magnitude forms of local pixels. | Complementary descriptors | Use histogram combination and minimize the Chi-square distance with free training. |
| Liao *et al.* [7] | Dominant LBP | Combination of local information (most frequently occurring pattern) and global information (Gabor feature). | Feature selection and learning | Less sensitive to noise and histogram equalization. A support vector machine (SVM) is used. |
| Guo *et al.* [8] | LBP variance | Integration rotation-variant LBP histogram and global matching. | Handling rotation | Use histogram combination and minimize the chi-square distance with free training. |
| Zhang *et al.* [10] | Monogenic-LBP | Supplement LBP with local phase and local surface measure with Riesz transform. | Handling rotation | 3-D normalized histogram and chi-square distance. |
| Guo *et al.* [22] | Adaptive LBP | Integration LBP with direction statistical features: mean, standard deviation, and adaptive coefficients. | Handling rotation | Chi-square distance is used to measure the histogram similarity. |
| Hafiane *et al.* [23] | Mean Binary Pattern | The mean value of local region is used as a threshold value. | Threshold and encoding | Capture contrast information of the local region and reduce the affect of scale changes by decomposing the images into smaller frequency ranges. |
| Yang and Wang [24] | LBP with Hamming distance | Integrate Hamming distance to LBP to reduce the error. | Feature selection and learning | Combine nonuniform pattern into uniform pattern by minimizing the Hamming distance. |
| Disparity map generation | | | | |
| Kanade [12] | SAD | Measure the absolute different intensity between the left and right pixels. | Local matching | Simplest similarity measurement. Sensitive to noise. |
| Bolles *et al.* [13] | NCC | Find the disparity that reduce the associated error and increase the similarity. | Local matching | More complex operation than SAD. Sensitive to noise. |
| Zabih and Woodfill [14] | Census | Nonparametric local transform, reducing the effect of variation due to camera. | Local matching | Improved the matching result near object boundaries but still sensitive to brightness difference. |
| Ohta and Kanade [15] | DP | Horizontal and vertical matching with intrascanline and interscanline. | Global matching | 3-D search space for long connected edges and 2-D for isolated edges. Sensitive to noise. |
| Boykov *et al.* [17] | Graphcut | Integration graph cut with two algorithms, $\alpha$–$\beta$ swap and $\alpha$-expansion to compute the local minimum. | Global matching | Assume two pixels in left and right images should have the same or similar intensity. Less sensitive to noise. |
| Felzenszwalb and Huttenlocher [19] | BP | Maximum product algorithm is used by transmitting a message in four directions. Multi-scale algorithm is used to reduce the processing time. | Global matching | Distance transform is used to minimize message cost and compute message in linear time. Less sensitive to noise. |
| Hirschmuller [20] | SGM-MI | Integration global methods with local stereo methods for pixel-wise matching. | Global matching | Mutual information is used for estimating matching cost. Robust to radiometric noise. |
| Heo *et al.* [21] | Adaptive NCC | Integration adaptive normal cross correlation with color formation model. | Global matching | Bilateral filter cost is substituted for window mean value cost. Insensitive to radiometric variation but still sensitive to multiple illumination conditions. |

Decision

# Summary Research Paper

1. Begin your summary with a brief introduction that includes the paper's title, authors, and publication information.

2. In a few sentences, introduce the research question or problem the paper addresses

3. Provide a concise overview of the methodology or approach used in the study, including the data sources, data collection and analysis methods, and any statistical techniques

4. Highlight the primary findings or results of the research. Use clear and straightforward language to communicate these findings. If the paper includes figures, tables, or graphs, refer to them as needed.

5. Conclude your summary with a closing statement that provides an overall assessment of the paper, such as its significance, relevance, or potential for future research

# Things You Need to Remember

Remember, the related work section should provide a solid foundation for your research, demonstrating that you are aware of existing work in your field and positioning your study in the context of that work.

# How to Write Related Work

**Step 1:** Define the Scope for finding related research papers

Clearly define the scope and objectives of your literature review. What specific topic or research question are you addressing? This will help you identify the key themes and relevant studies to include.

Robust Stereo Matching

Stereo Matching with CNN

Stereo Matching Survey

Stereo Matching Adverse Driving Condition

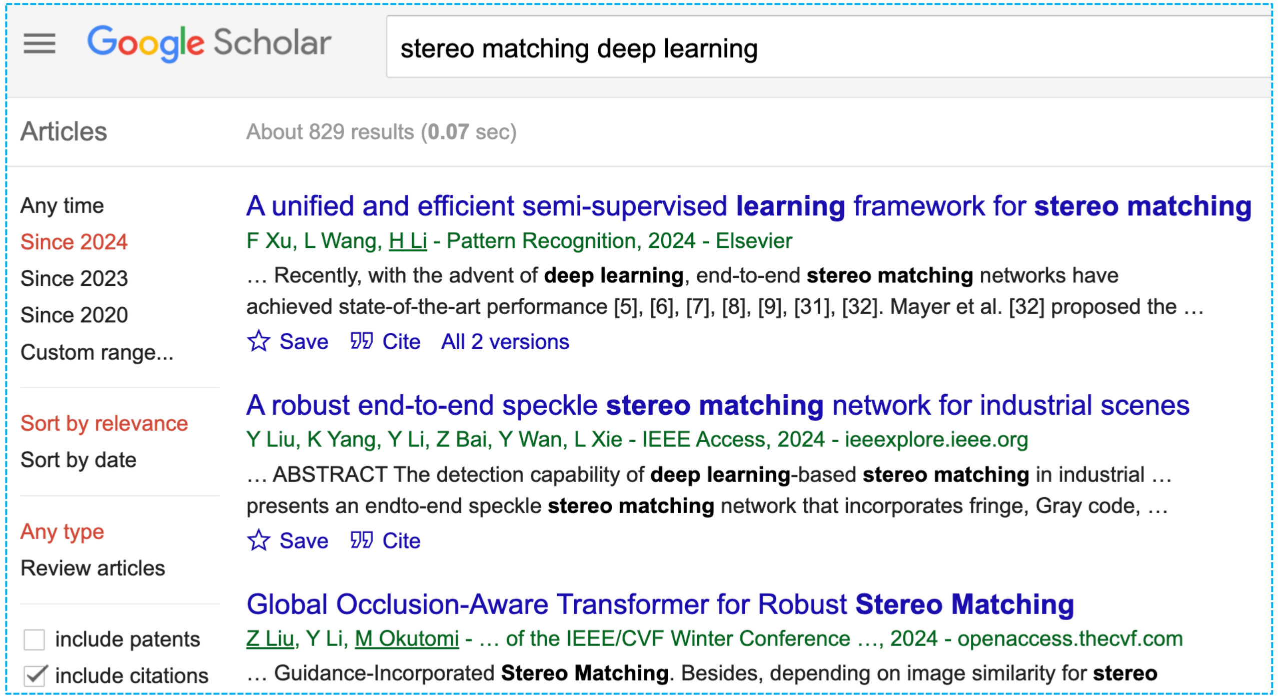
Stereo Matching Transformer

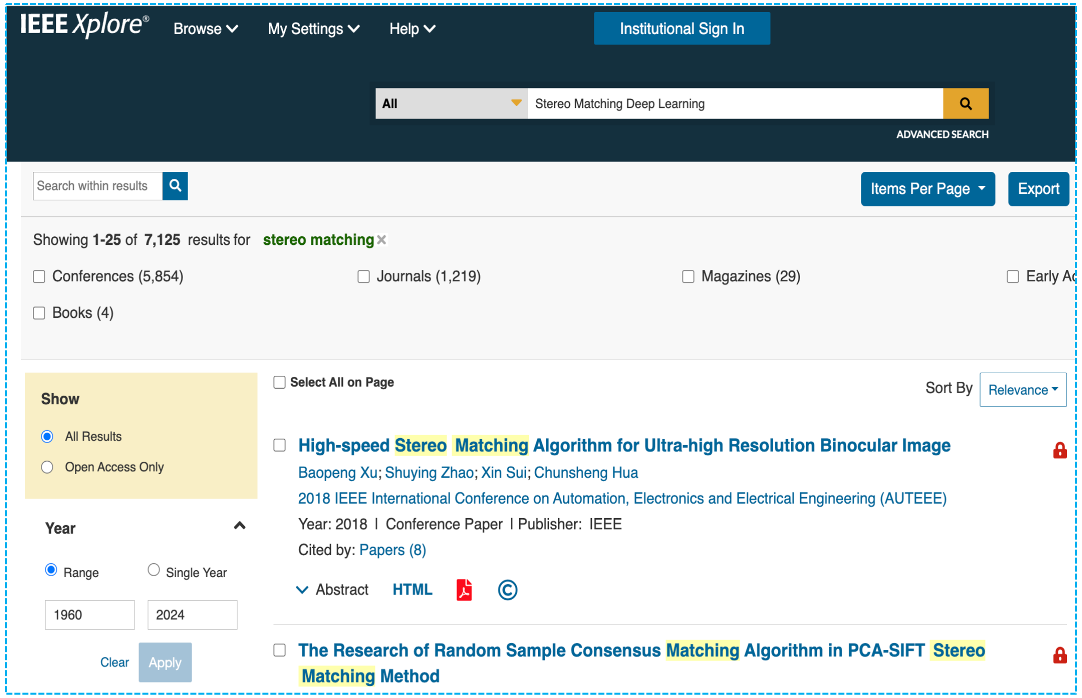
Stereo Matching using Deep Learning

# How to Write Related Work

## Step 2: Search for Relevant Literature

Conduct a comprehensive literature search using academic databases, journals, books, and other reputable sources. Use relevant keywords and Boolean operators to refine your search.



# How to Write Related Work

## Step 3: Organize the Literature

Organize the literature based on themes, topics, or methodologies. This can be done chronologically, conceptually, or by grouping studies that share common characteristics.

| | Year | Type | Purpose |
|---|---|---|---|
| Make3D [14] | 2009 | Real | Monocular depth |
| KITTI2012 [15] | 2012 | Real | Stereo |
| MPI Sintel [16] | 2012 | Synthetic | Optical flow |
| NYU2 [17] | 2012 | Real - indoor | Monocular depth, object segmentation |
| RGB-D SLAM [18] | 2012 | Real | SLAM |
| SUN3D [19] | 2013 | Real - rooms | Monocular video |
| Middleburry [20] | 2014 | Indoor | Stereo |
| KITTI 2015 [21] | 2015 | Real | Stereo |
| KITTI-MVS2015 [21] | 2015 | Real | MVS |
| FlyingThings3D, Monkaa, Driving [22] | 2016 | Synthetic | Stereo, Video, Optical flow |
| CityScapes [23] | 2016 | Street scenes | Semantic seg., dense labels |
| | | | Semantic seg.,coarse labels |
| DTU [24] | 2016 | Real, small objects | MVS |
| ETH3D [25] | 2017 | Real, in/outdoor | Low-res, Stereo |
| | | | Low-res, MVS on video |
| | | | High-res, MVS on images from DSLR camera |
| SUNCG [26] | 2017 | Synthetic, indoor | Scene completion |
| MVS-Synth [27] | 2018 | Synth - urban | MVS |
| MegaDepth [28] | 2018 | Real (Internet images) | Monocular, Eucl. and ord. depth |
| Jeon and Lee [29] | 2018 | Real | Depth enhancement |
| OmniThings [30], [31] | 2019 | Synthetic, fish-eye images | Omnidirectional MVS |
| OmniHouse [30], [31] | 2019 | Synthetic, fish-eye images | Omnidirectional MVS |
| HR-VS [32] | 2019 | Synthetic, outdoor | High res. stereo |
| | | Real, outdoor | High res. stereo |
| DrivingStereo [33] | 2019 | Driving | High res. stereo |
| ApolloScape [34] | 2019 | Auto. driving | High res. stereo |
| A2D2 [35] | 2020 | Auto. driving | High res. stereo |

| Method | Year | Feature computation | |
|---|---|---|---|
| | | Architectures | Dimension |
| Zagoruyko [37] | 2015 | ConvNet | Multiscale |
| Han [38] | 2015 | ConvNet | Fixed scale |
| Zbontar [39] | 2015 | ConvNet | Fixed scale |
| Chen [40] | 2015 | ConvNet | Multiscale |
| Simo [41] | 2015 | ConvNet | Fixed scale |
| Zbontar [42] | 2016 | ConvNet | Fixed scale |
| Balantas [43] | 2016 | ConvNet | Fixe scale |
| Mayer [22] | 2016 | ConvNet | Fixed-scale |
| Luo [44] | 2016 | ConvNet | Fixed scale |
| Kumar [45] | 2016 | ConvNet | Fixed scale |
| Shaked [46] | 2017 | Highway network with multilevel skip connections | Fixed scale |
| Hartmann [47] | 2017 | ConvNet | Fixed scale |
| Park [48] | 2017 | ConvNet | Fixed scale |
| Ye [49] | 2017 | ConvNet Multisize pooling | Fixed scale |

AI VIETNAM
All-in-One Course

# How to Write Related Work

**Step 3:** Organize the Literature

Group related works based on common themes, topics, or methodologies. This helps in presenting a coherent and structured review.

## 2. Related Works

**Deep Stereo Matching.** Recently, the success of convolution neural networks has driven the community to develop learning based solutions for stereo matching [48, 26, 28, 22, 5, 18, 15, 24, 43, 21]. Specifically, Mayer *et al.* [26] proposed the first end-to-end method DispNetC, which directly calculated the correlation between left and right features by multiplying the pixels at the corresponding position. Chang *et al.* introduced PSMNet [5], using a spatial pyramid pooling module to leverage the capacity of global context information in different scales. Based on this, Guo *et al.* [15] proposed GwcNet via group-wise correlation, achieving better performance and reducing parameters simultaneously. For most methods, the diversity in disparity distribution is the main challenge for model performance, which can be improved through an iterative mechanism. Following the great success of RAFT [39] in optical flow task, RaftStereo [24] was proposed for stereo matching with iterative refinement. Li *et al.* [21] proposed CREStereo, which illustrates the effectiveness of cascaded recurrent network.

**Robust Stereo Matching.** Robust stereo matching oriented toward robustness and real-world applications is a less explored problem. Jia *et al.* [41] introduced an end-to-end network with scene geometry priors to improve the network's generalization ability to unseen scenes. Song *et al.* [36] introduced a domain adaptation method to handle the gap between synthetic and real-world domains. Zhang *et al.* [50] proposed a domain-invariant approach via a domain normalization layer and learnable graph-based filter. MCV-MFC [23] proposed a two-stage training strategy to transfer the model to target datasets gently. Shen *et al.* [35] proposed CFNet, a cascaded and fused cost volume based network to deal with the domain difference, illustrating the potential of cascaded architecture for robust vision tasks. However, it still suffers from a lack of flexibility for modeling sampling in complicated structures.

**Real-time Stereo Matching.** Several recent works [19, 11, 3, 42, 44] focus on real-time performance while maintaining satisfactory accuracy. StereoNet [19] introduced an edge-preserving refinement network to leverage left images to recover high frequency details towards information loss at low resolution. DeepPruner [11] built a sparse cost volume by PatchMatch [3], and pruned the search space based on the predicted disparities, which were further refined under the guidance of image features. Xu *et al.* [45] proposed AANet, designed a sparse points based cost aggregation method and replaced the commonly used 3D convolutions to achieve fast inference speed. Xu *et al.* [44] introduced Fast-ACVNet, which adopted an attention mechanism to suppress redundant information and enhance matching-related information in the concatenation volume, which is quite efficient. In this paper, we also introduce a real-time stereo matching network based on the proposed network architecture while maintaining accuracy.

Deep Stereo Matching

Robust Stereo Matching

Real-time Stereo Matching

https://arxiv.org/pdf/2307.14071.pdf

# How to Write Related Work

**Step 3:** Organize the Literature

Group related works based on common themes, topics, or methodologies. This helps in presenting a coherent and structured review.

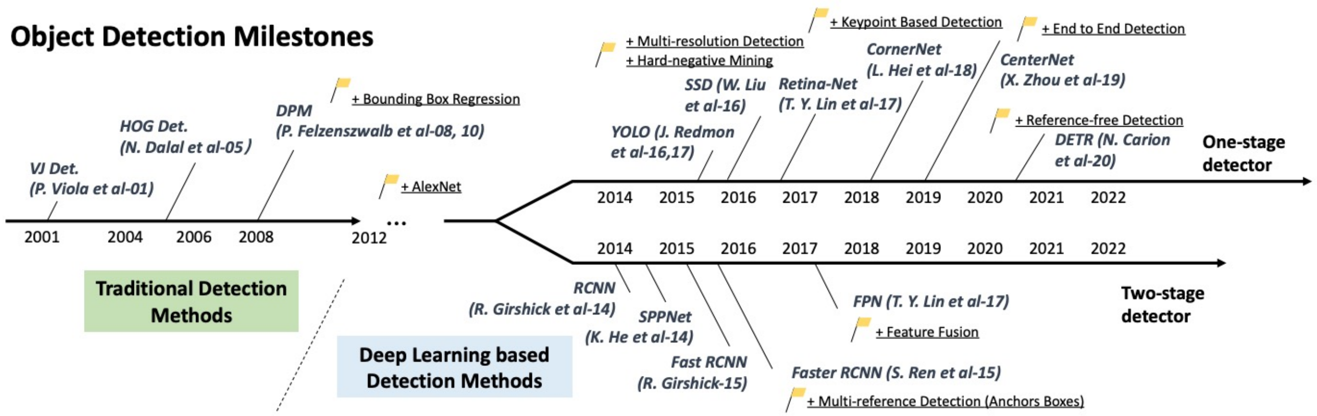



Fig. 2: A road map of object detection. Milestone detectors in this figure: VJ Det. [10, 11], HOG Det. [12], DPM [13–15], RCNN [16], SPPNet [17], Fast RCNN [18], Faster RCNN [19], YOLO [20–22], SSD [23], FPN [24], Retina-Net [25], CornerNet [26], CenterNet [27], DETR [28].

- Traditional Detection Methods
- Deep Learning Detection Methods
- Transformer Detection Methods

https://arxiv.org/pdf/2307.14071.pdf

AI VIETNAM
All-in-One Course

# How to Write Related Work

## Step 4: Start Broad, Then Narrow

Begin with a broad overview of the field or topic, providing context for your research. Gradually narrow down to the specific aspects that are closely related to your work.

[3]. In practice, however, almost all existing stereo matching algorithms do not take into consideration this color-constancy process. Therefore, there is a pressing need to include this process into the matching mechanism.

In this paper, we present a new stereo matching measure that is robust in handling various radiometric changes, including local radiometric variations caused by varying lighting geometry, as well as global radiometric variations that are brought about by changes in illuminant color and camera parameter [4]. It should be noted that this paper is based on the assumption of the Lambertian world. Under the assumption, the color formation process was explicitly modeled, unlike other methods. Then, we have extracted the invariant color information from the color model and thereby propose, in this paper, a new matching measure, called *Adaptive Normalized Cross-Correlation (ANCC)*, which is robust to various radiometric changes.

### 1.2 Related Works

Hirschmüller and Scharstein carried out an evaluation of various cost functions for stereo matching on radiometrically different images caused by factors such as light configuration, camera exposure, gamma correction variations, and noise, etc [5]. Under various radiometric conditions, they compared Birchfield and Tomasi data cost (BT) [6], BT with Laplacian of Gaussian (LoG) filtering [7], BT with Mean filtering, BT with Rank transform [8], Normalized Cross-Correlation (NCC), and Hierarchical Mutual Information (HMI) [9] with correlation-based method, as well as the semiglobal and global method. Although BT cost is a very popular data cost which is insensitive to camera sampling, the major drawback of it is that it cannot handle radiometrically different images as it employs the use of linearly interpolated function of intensity values. Although the LoG filter calculates second-order derivatives that are insensitive to outliers [7], BT with LoG filtering itself is insufficient for use in stereo matching. BT with Rank transform [8] possesses a robust property to radiometric variations because the Rank transform is based on the principle that the rank between pixel intensities does not vary after radiometric variations. Although it is effective in handling global radiometric changes, it is weak with respect to local radiometric variations. NCC is a highly popular as well as traditional measure for matching contrast-varying images [10], [11]. It measures only the cosine of an angle between matching vectors because normalization renders the matching vectors to have zero mean and one standard deviation. Therefore, NCC is suitable only for matching affine-transformed intensity or color values and it also suffers from the fattening effect that object boundaries are not reconstructed correctly similar to the Sum of Absolute Difference (SAD) and the Sum of Squared Difference (SSD). Mutual Information (MI) has been used as a similarity measure in computer vision [9], [12], [13], [14]. For stereo matching, Kim et al. [14] proposed a pixelwise data cost in the MAP-MRF framework based on mutual information. Their work was improved by Hirschmüller [9] and extended to handle occlusion and speed up the computation by using the Hierarchical Mutual Information (HMI). The advantage of the MI-based cost is that it enables finding the correspondence between globally transformed image pairs by utilizing the joint histogram of intensities between two input images. However, local radiometric variations due to lighting geometry changes cannot be handled by MI-based cost. Because of the just-discussed reasons, Hirschmüller and Scharstein [5] concluded that all compared costs previously described fail to achieve success with respect to strong local radiometric changes caused by the light configuration changes. Besides, Wang et al. [15] presented a new invariant measure called light transport constancy (LTC) based on a rank constraint for non-Lambertian surfaces. Their method required at least two stereo image pairs with different illumination conditions to be available for making use of rank constraint. Ogale and Aloimonos [16] presented a contrast-robust stereo matching algorithm for local matching using multiple frequency channels. Negahdaripour [17] proposed a general linear brightness constraint for handling radiometric variations between images. The principle behind his model was that the linear transformation of the brightness between image patches was assumed, and the linear transformation parameters and the optical flow vectors were simultaneously estimated. Zhang et al. [18] presented a unified MRF framework for the estimation of illumination variation and disparity map simultaneously. Their framework sought to define Illumination Ratio Map (IRM), and assumed disparity and illumination smoothness.

A major drawback of most of the above-mentioned methods is that these utilize only raw intensity (or color) information, which can be sensitive to severe radiometric variations for stereo matching due to the fact that raw intensity depends only on the direction of the light and the surface normal and does not take into consideration the

---

We comprehensively evaluate our method on the PASCAL VOC detection benchmarks [11] where RPNs with Fast R-CNNs produce detection accuracy better than the strong baseline of Selective Search with Fast R-CNNs. Meanwhile, our method waives nearly all computational burdens of Selective Search at test-time—the effective running time for proposals is just 10 milliseconds. Using the expensive very deep models of [3], our detection method still has a frame rate of 5fps (*including all steps*) on a GPU, and thus is a practical object detection system in terms of both speed and accuracy. We also report results on the MS COCO dataset [12] and investigate the improvements on PASCAL VOC using the COCO data. Code has been made publicly available at `https://github.com/shaoqingren/faster_rcnn` (in MATLAB) and `https://github.com/rbgirshick/py-faster-rcnn` (in Python).

A preliminary version of this manuscript was published previously [10]. Since then, the frameworks of RPN and Faster R-CNN have been adopted and generalized to other methods, such as 3D object detection [13], part-based detection [14], instance segmentation [15], and image captioning [16]. Our fast and effective object detection system has also been built in com-

1. Since the publication of the conference version of this paper [10], we have also found that RPNs can be trained jointly with Fast R-CNN networks leading to less training time.

### 2 RELATED WORK

**Object Proposals.** There is a large literature on object proposal methods. Comprehensive surveys and comparisons of object proposal methods can be found in [19], [20], [21]. Widely used object proposal methods include those based on grouping super-pixels (*e.g.*, Selective Search [4], CPMC [22], MCG [23]) and those based on sliding windows (*e.g.*, objectness in windows [24], EdgeBoxes [6]). Object proposal methods were adopted as external modules independent of the detectors (*e.g.*, Selective Search [4] object detectors, R-CNN [5], and Fast R-CNN [2]).

**Deep Networks for Object Detection.** The R-CNN method [5] trains CNNs end-to-end to classify the proposal regions into object categories or background. R-CNN mainly plays as a classifier, and it does not predict object bounds (except for refining by bounding box regression). Its accuracy depends on the performance of the region proposal module (see comparisons in [20]). Several papers have proposed ways of using deep networks for predicting object bounding boxes [25], [9], [26], [27]. In the OverFeat method [9], a fully-connected layer is trained to predict the box coordinates for the localization task that assumes a single object. The fully-connected layer is then turned

2. http://image-net.org/challenges/LSVRC/2015/results

# How to Write Related Work

## Step 5: Critical Evaluation

Don't just summarize each paper; critically evaluate them. Discuss the strengths and weaknesses of each study and highlight any gaps in the existing literature that your research aims to fill.

We comprehensively evaluate our method on the PASCAL VOC detection benchmarks [11] where RPNs with Fast R-CNNs produce detection accuracy better than the strong baseline of Selective Search with Fast R-CNNs. Meanwhile, our method waives nearly all computational burdens of Selective Search at test-time—the effective running time for proposals is just 10 milliseconds. Using the expensive very deep models of [3], our detection method still has a frame rate of 5fps (*including all steps*) on a GPU, and thus is a practical object detection system in terms of both speed and accuracy. We also report results on the MS COCO dataset [12] and investigate the improvements on PASCAL VOC using the COCO data. Code has been made publicly available at https://github.com/shaoqingren/faster_rcnn (in MATLAB) and https://github.com/rbgirshick/py-faster-rcnn (in Python).

A preliminary version of this manuscript was published previously [10]. Since then, the frameworks of RPN and Faster R-CNN have been adopted and generalized to other methods, such as 3D object detection [13], part-based detection [14], instance segmentation [15], and image captioning [16]. Our fast and effective object detection system has also been built in com-

1. Since the publication of the conference version of this paper [10], we have also found that RPNs can be trained jointly with Fast R-CNN networks leading to less training time.

### 2 Related Work

**Object Proposals.** There is a large literature on object proposal methods. Comprehensive surveys and comparisons of object proposal methods can be found in [19], [20], [21]. Widely used object proposal methods include those based on grouping super-pixels (*e.g.*, Selective Search [4], CPMC [22], MCG [23]) and those based on sliding windows (*e.g.*, objectness in windows [24], EdgeBoxes [6]). Object proposal methods were adopted as external modules independent of the detectors (*e.g.*, Selective Search [4] object detectors, R-CNN [5], and Fast R-CNN [2]).

**Deep Networks for Object Detection.** The R-CNN method [5] trains CNNs end-to-end to classify the proposal regions into object categories or background. R-CNN mainly plays as a classifier, and it does not predict object bounds (except for refining by bounding box regression). Its accuracy depends on the performance of the region proposal module (see comparisons in [20]). Several papers have proposed ways of using deep networks for predicting object bounding boxes [25], [9], [26], [27]. In the OverFeat method [9], a fully-connected layer is trained to predict the box coordinates for the localization task that assumes a single object. The fully-connected layer is then turned

2. http://image-net.org/challenges/LSVRC/2015/results

### 2 Related work

Our work build on prior work in several domains: bipartite matching losses for set prediction, encoder-decoder architectures based on the transformer, parallel decoding, and object detection methods.

#### 2.1 Set Prediction

There is no canonical deep learning model to directly predict sets. The basic set prediction task is multilabel classification (see e.g., [40,33] for references in the context of computer vision) for which the baseline approach, one-vs-rest, does not apply to problems such as detection where there is an underlying structure between elements (i.e., near-identical boxes). The first difficulty in these tasks is to avoid near-duplicates. Most current detectors use postprocessings such as non-maximal suppression to address this issue, but direct set prediction are postprocessing-free. They need global inference schemes that model interactions between all predicted elements to avoid redundancy. For constant-size set prediction, dense fully connected networks [9] are sufficient but costly. A general approach is to use auto-regressive sequence models such as recurrent neural networks [48]. In all cases, the loss function should be invariant by a permutation of the predictions. The usual solution is to design a loss based on the Hungarian algorithm [20], to find a bipartite matching between ground-truth and prediction. This enforces permutation-invariance, and guarantees that each target element has a unique match. We follow the bipartite matching loss approach. In contrast to most prior work however, we step away from autoregressive models and use transformers with parallel decoding, which we describe below.

#### 2.2 Transformers and Parallel Decoding

Transformers were introduced by Vaswani *et al.* [47] as a new attention-based building block for machine translation. Attention mechanisms [2] are neural network layers that aggregate information from the entire input sequence. Transformers introduced self-attention layers, which, similarly to Non-Local Neural Networks [49], scan through each element of a sequence and update it by aggregating information from the whole sequence. One of the main advantages of attention-based models is their global computations and perfect memory, which makes them more suitable than RNNs on long sequences. Transformers are now

AI VIETNAM
All-in-One Course

# How to Write Related Work

## Step 6: Show Progression

Demonstrate how the field has progressed over time and how your research fits into this progression. This could involve discussing seminal studies that laid the groundwork for current research.

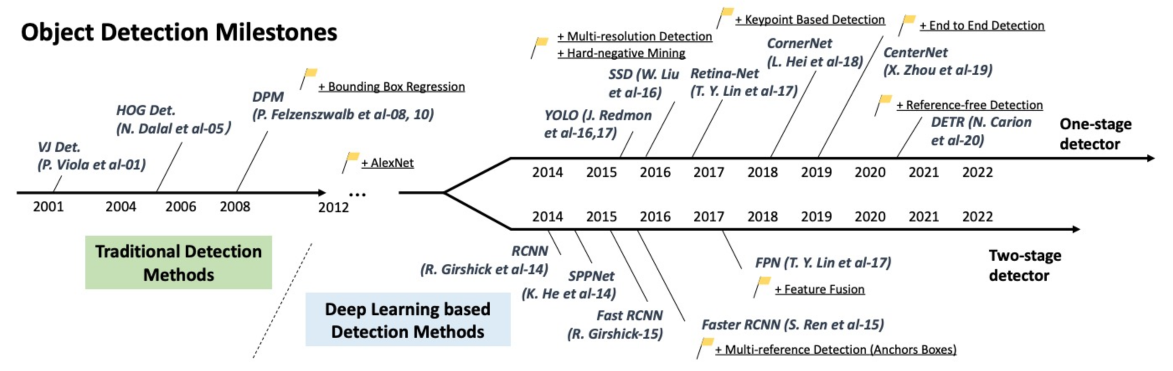



Fig. 2: A road map of object detection. Milestone detectors in this figure: VJ Det. [10, 11], HOG Det. [12], DPM [13–15], RCNN [16], SPPNet [17], Fast RCNN [18], Faster RCNN [19], YOLO [20–22], SSD [23], FPN [24], Retina-Net [25], CornerNet [26], CenterNet [27], DETR [28].

*1) Milestones: Traditional Detectors:* If we consider today's object detection technique as a revolution driven by deep learning, then back in the 1990s, we would see the ingenious design and long-term perspective of early computer vision. Most of the early object detection algorithms were built based on handcrafted features. Due to the lack of effective image representation at that time, people have to design sophisticated feature representations and a variety of speed-up skills.

**Viola Jones Detectors:** In 2001, P. Viola and M. Jones achieved real-time detection of human faces for the first time without any constraints (e.g., skin color segmentation) [10, 11]. Running on a 700MHz Pentium III CPU, the detector was tens or even hundreds of times faster than other algorithms in its time under comparable detection accuracy. The VJ detector follows a most straightforward way of detection, i.e., sliding windows: to go through all possible locations and scales in an image to see if any window contains a human face. Although it seems to be a very simple process, the calculation behind it was far beyond the computer's power of its time. The VJ detector has dramatically improved its detection speed by incorporating three important techniques: "integral image", "feature selection", and "detection cascades" (to be introduced in section III).

**HOG Detector:** In 2005, N. Dalal and B. Triggs proposed Histogram of Oriented Gradients (HOG) feature descriptor [12]. HOG can be considered as an important improvement of the scale-invariant feature transform [29, 30] and shape contexts [31] of its time. To balance the feature invariance (including translation, scale, illumination, etc) and the nonlinearity, the HOG descriptor is designed to be computed on a

**Deformable Part-based Model (DPM):** DPM, as the winners of VOC-07, -08, and -09 detection challenges, was the epitome of the traditional object detection methods. DPM was originally proposed by P. Felzenszwalb [13] in 2008 as an extension of the HOG detector. It follows the detection philosophy of "divide and conquer", where the training can be simply considered as the learning of a proper way of decomposing an object, and the inference can be considered as an ensemble of detections on different object parts. For example, the problem of detecting a "car" can be decomposed to the detection of its window, body, and wheels. This part of the work, a.k.a. "star-model", was introduced by P. Felzenszwalb *et al.* [13]. Later on, R. Girshick has further extended the star model to the "mixture models" to deal with the objects in the real world under more significant variations and has made a series of other improvements [14, 15, 33, 34].

Although today's object detectors have far surpassed DPM in detection accuracy, many of them are still deeply influenced by its valuable insights, e.g., mixture models, hard negative mining, bounding box regression, context priming, etc. In 2010, P. Felzenszwalb and R. Girshick were awarded the "lifetime achievement" by PASCAL VOC.

*2) Milestones: CNN based Two-stage Detectors:* As the performance of hand-crafted features became saturated, the research of object detection reached a plateau after 2010. In 2012, the world saw the rebirth of convolutional neural networks [35]. As a deep convolutional network is able to learn robust and high-level feature representations of an image, a natural question arises: can we introduce it to object detection? R. Girshick *et al.* took the lead to break the deadlocks in

# How to Write Related Work

## Step 7: Synthesize Information

Synthesize the information from different sources to provide a comprehensive overview. Show how the existing works contribute to the current state of knowledge and identify any consensus or debates in the literature.

### 2.3 Object detection

Most modern object detection methods make predictions relative to some initial guesses. Two-stage detectors [37,5] predict boxes w.r.t. proposals, whereas single-stage methods make predictions w.r.t. anchors [23] or a grid of possible object centers [53,46]. Recent work [52] demonstrate that the final performance of these systems heavily depends on the exact way these initial guesses are set. In our model we are able to remove this hand-crafted process and streamline the detection process by directly predicting the set of detections with absolute box prediction w.r.t. the input image rather than an anchor.

**Set-based loss.** Several object detectors [9,25,35] used the bipartite matching loss. However, in these early deep learning models, the relation between different prediction was modeled with convolutional or fully-connected layers only and a hand-designed NMS post-processing can improve their performance. More recent detectors [37,23,53] use non-unique assignment rules between ground truth and predictions together with an NMS.

Learnable NMS methods [16,4] and relation networks [17] explicitly model relations between different predictions with attention. Using direct set losses, they do not require any post-processing steps. However, these methods employ additional hand-crafted context features like proposal box coordinates to model relations between detections efficiently, while we look for solutions that reduce the prior knowledge encoded in the model.

# How to Write Related Work

## Step 8: Connect to Your Research

Clearly articulate how each work is related to your research. Highlight the aspects that are directly relevant to your study and explain how they inform your research questions, objectives, or methodology.

### 2.3 Object detection

Most modern object detection methods make predictions relative to some initial guesses. Two-stage detectors [37,5] predict boxes w.r.t. proposals, whereas single-stage methods make predictions w.r.t. anchors [23] or a grid of possible object centers [53,46]. Recent work [52] demonstrate that the final performance of these systems heavily depends on the exact way these initial guesses are set. In our model we are able to remove this hand-crafted process and streamline the detection process by directly predicting the set of detections with absolute box prediction w.r.t. the input image rather than an anchor.

**Set-based loss.** Several object detectors [9,25,35] used the bipartite matching loss. However, in these early deep learning models, the relation between different prediction was modeled with convolutional or fully-connected layers only and a hand-designed NMS post-processing can improve their performance. More recent detectors [37,23,53] use non-unique assignment rules between ground truth and predictions together with an NMS.

Learnable NMS methods [16,4] and relation networks [17] explicitly model relations between different predictions with attention. Using direct set losses, they do not require any post-processing steps. However, these methods employ additional hand-crafted context features like proposal box coordinates to model relations between detections efficiently, while we look for solutions that reduce the prior knowledge encoded in the model.

**Recurrent detectors.** Closest to our approach are end-to-end set predictions for object detection [43] and instance segmentation [41,30,36,42]. Similarly to us, they use bipartite-matching losses with encoder-decoder architectures based on CNN activations to directly produce a set of bounding boxes. These approaches, however, were only evaluated on small datasets and not against modern baselines. In particular, they are based on autoregressive models (more precisely RNNs), so they do not leverage the recent transformers with parallel decoding.

# How to Write Related Work

## Step 9: Cite Key References

Ensure that you cite the most relevant and influential works in your field. This demonstrates a thorough understanding of the existing literature and establishes the foundation for your own research.

**References**

1. Al-Rfou, R., Choe, D., Constant, N., Guo, M., Jones, L.: Character-level language modeling with deeper self-attention. In: AAAI Conference on Artificial Intelligence (2019)
2. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: ICLR (2015)
3. Bello, I., Zoph, B., Vaswani, A., Shlens, J., Le, Q.V.: Attention augmented convolutional networks. In: ICCV (2019)
4. Bodla, N., Singh, B., Chellappa, R., Davis, L.S.: Soft-NMS improving object detection with one line of code. In: ICCV (2017)
5. Cai, Z., Vasconcelos, N.: Cascade R-CNN: High quality object detection and instance segmentation. PAMI (2019)
6. Chan, W., Saharia, C., Hinton, G., Norouzi, M., Jaitly, N.: Imputer: Sequence modelling via imputation and dynamic programming. arXiv:2002.08926 (2020)
7. Cordonnier, J.B., Loukas, A., Jaggi, M.: On the relationship between self-attention and convolutional layers. In: ICLR (2020)
8. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (2019)
9. Erhan, D., Szegedy, C., Toshev, A., Anguelov, D.: Scalable object detection using deep neural networks. In: CVPR (2014)
10. Ghazvininejad, M., Levy, O., Liu, Y., Zettlemoyer, L.: Mask-predict: Parallel decoding of conditional masked language models. arXiv:1904.09324 (2019)
11. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: AISTATS (2010)

Remember to cite several related research studies published in the journal to which you plan to submit your work

**References**

[1] Joydeep Biswas and Manuela Veloso. Depth camera based localization and navigation for indoor mobile robots. In *RGB-D Workshop at RSS*, volume 2011, page 21, 2011. 1

[2] Jia-Ren Chang and Yong-Sheng Chen. Pyramid stereo matching network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5410–5418, 2018. 8

[3] Chenyi Chen, Ari Seff, Alain Kornhauser, and Jianxiong Xiao. Deepdriving: Learning affordance for direct perception in autonomous driving. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2722–2730, 2015. 1

[4] Shuo Cheng, Zexiang Xu, Shilin Zhu, Zhuwen Li, Li Erran Li, Ravi Ramamoorthi, and Hao Su. Deep stereo using adaptive thin volume representation with uncertainty awareness. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2524–2534, 2020. 3, 4, 5

[5] Xinjing Cheng, Peng Wang, and Ruigang Yang. Learning depth with convolutional spatial propagation network. In *IEEE Transactions on pattern analysis and machine intelligence (TPAMI)*, 2019. 3

[6] Xuelian Cheng, Yiran Zhong, Mehrtash Harandi, Yuchao Dai, Xiaojun Chang, Hongdong Li, Tom Drummond, and Zongyuan Ge. Hierarchical neural architecture search for deep stereo matching. *Advances in Neural Information Processing Systems (NIPS)*, 33, 2020. 8

*IEEE transactions on pattern analysis and machine intelligence(TPAMI)*, 42(10):2410–2422, 2019. 8

[14] Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry, Ryan Kennedy, Abraham Bachrach, and Adam Bry. End-to-end learning of geometry and context for deep stereo regression. In *IEEE International Conference on Computer Vision (ICCV)*, pages 66–75, 2017. 3, 5

[15] Zhengfa Liang, Yiliu Feng, Yulan Guo, Hengzhu Liu, Wei Chen, Linbo Qiao, Li Zhou, and Jianfeng Zhang. Learning for disparity estimation through feature constancy. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2811–2820, 2018. 8

[16] Zhengfa Liang, Yulan Guo, Yiliu Feng, Wei Chen, Linbo Qiao, Li Zhou, Jianfeng Zhang, and Hengzhu Liu. Stereo matching using multi-level cost volume and multi-scale feature constancy. In *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*, 2019. 3, 6, 8

[17] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. In *Advances in neural information processing systems (NIPS)*, pages 4898–4906, 2016. 2

[18] Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4040–4048, 2016. 3

# How Do Reviewers Expect

### Comprehensiveness

Reviewers look for evidence that the author has conducted a thorough literature review. This includes covering seminal works, recent publications, and key research in the field. A comprehensive review demonstrates the author's familiarity with existing literature.

### Critical Analysis

Reviewers expect more than just a list of references. They look for a critical analysis of the existing literature, highlighting strengths, weaknesses, and limitations of prior work. This analysis helps in positioning the author's research in the context of what is already known.

### Integration with the Research

Reviewers assess how well the related work section aligns with the research question or problem statement. The section should build a logical bridge between what is already known and the proposed research.

### Relevance

The related work section should focus on studies directly related to the research topic. Reviewers assess whether the cited works are relevant to the research question or problem addressed in the paper.

### Identification of Gaps

A good related work section identifies gaps or limitations in existing research that the current study aims to address. Reviewers assess whether the author convincingly argues for the need for their research based on identified

### Up-to-Date References

The inclusion of recent and up-to-date references is essential. Reviewers check to see if the author has considered the most current research in the field.

# Reviewers' Comments

Reviewer Comment: "The related work section seems limited in scope. Consider expanding your literature review to include more recent and relevant studies in the field."

Reviewer Comment: "Several seminal works are missing from your related work section. Please include references to [specific papers] as they are critical to the understanding of the background."

Reviewer Comment: "The related work needs more critical analysis. Rather than just summarizing each work, provide insights into how these studies contribute or fall short in addressing the research gap."

Reviewer Comment: "Several references in the related work section are outdated. Please update the literature review with more recent studies to reflect the current state of the field."

# Reviewers' Comments

Reviewer Comment: "The organization of the related work section is confusing. Consider reordering the studies or grouping them by thematic relevance to improve clarity for the reader."

Reviewer Comment: "The related work section lacks a clear identification of research gaps. Please discuss the limitations of existing studies and how your research fills these gaps."

Reviewer Comment: "The related work appears biased towards a particular perspective. Ensure that you present a balanced view by including studies that may have different findings or viewpoints."

Reviewer Comment: "There is some redundancy between the introduction and related work sections. Clarify the distinct purpose of each section to avoid repetition."

Reviewer Comment: "The analysis of the related work is quite surface-level. Provide a more in-depth examination of methodologies, results, and implications of the cited studies."

# Outline

- How to Write Related Work Section
- How to Write Introduction Section
- Assignment

# How to Write an Introduction

<<Paper' title>>

Abstract

Keyword

1. Introduction

2. Related Work

3. Proposed Method

4. Experimental Results

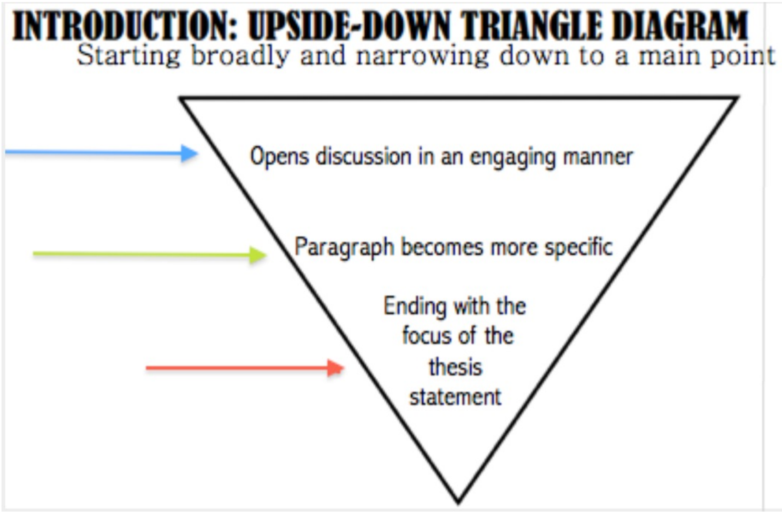
5. Conclusion

References

# How to Write an Introduction

# How to Write an Introduction

## Q: How Does an Introduction Differ from a Literature Review?




**Answer:**

- The introduction introduces your research question(s), purpose, objectives, or hypotheses.
- The literature review critically evaluates the existing research in greater detail, summarizing and synthesizing important articles

# How to Write an Introduction

Q: How Does an Introduction Differ from a Literature Review?

- INTRODUCE YOUR TOPIC
- CREATE SOME CONTEXT AND BACKGROUND
- TELL YOUR READER ABOUT THE RESEARCH YOU PLAN TO CARRY OUT
- STATE YOUR RATIONALE
- EXPLAIN WHY YOUR RESEARCH IS IMPORTANT
- STATE YOUR HYPOTHESIS

**Answer:**

- The introduction introduces your research question(s), purpose, objectives, or hypotheses.
- The literature review critically evaluates the existing research in greater detail, summarizing and synthesizing important articles

# How to Write an Introduction

Q: How Does an Introduction Differ from a Literature Review?

- General Research Topic
- Your Research Topic
- Current Problems
- Solution and Hypotheses
- Introduction to the Method
- Main contributions

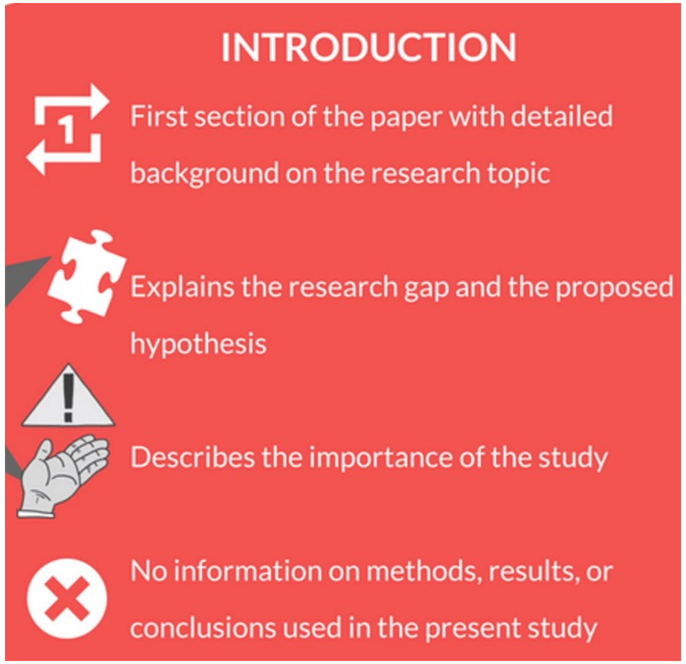
**Answer:**

- The introduction introduces your research question(s), purpose, objectives, or hypotheses.
- The literature review critically evaluates the existing research in greater detail, summarizing and synthesizing important articles

AI VIETNAM
All-in-One Course

# How to Write an Introduction

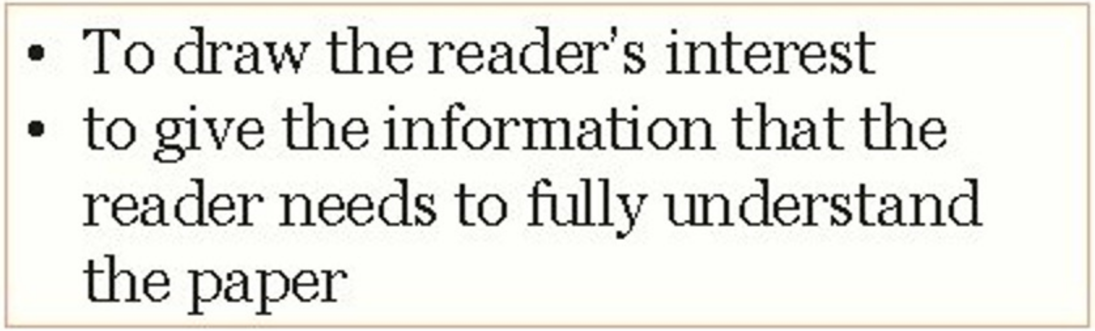
Q: How Does an Introduction Differ from a Literature Review?




**Answer:**
- The introduction introduces your research question(s), purpose, objectives, or hypotheses.
- The literature review critically evaluates the existing research in greater detail, summarizing and synthesizing important articles

# How to Write an Introduction

Q: How Does an Introduction Differ from a Literature Review?

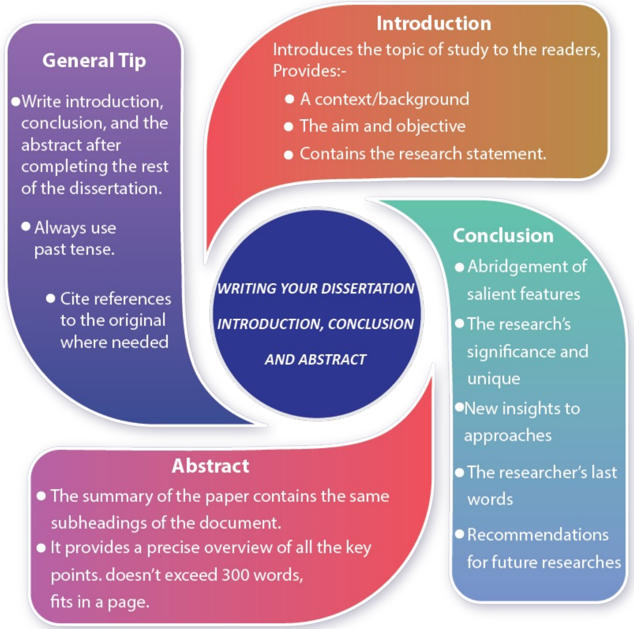

- To draw the reader's interest
- to give the information that the reader needs to fully understand the paper

**Answer:**

- The introduction introduces your research question(s), purpose, objectives, or hypotheses.
- The literature review critically evaluates the existing research in greater detail, summarizing and synthesizing important articles

# How to Write an Introduction

Q: How Does an Introduction Differ from a Literature Review?




**Answer:**

- The introduction introduces your research question(s), purpose, objectives, or hypotheses.
- The literature review critically evaluates the existing research in greater detail, summarizing and synthesizing important articles

*Abstract*—**The performance of stereo matching algorithms strongly depends on the quality of the stereo data/matching cost. Most state-of-the-art data costs require expert knowledge for the design of a transformation function, such as census for handling gray-level changes monotonically, adaptive normalized cross correlation for handling Lambertian cases, guided filtering for preserving edge information, and local density encoding for handling illumination differences. However, it is difficult to design a complex transformation function to handle unknown factors that often occur in driving conditions such as snow, rain, and sun. Therefore, this paper has investigated the deep learning strategy to develop a novel stereo matching cost model without using much expert knowledge. Experimental results show that the proposed deep learning model obtains better results than the state-of-the-art stereo matching cost as judged by the standard KITTI benchmark, Middlebury, and HCI datasets.**

*Index Terms*—**Stereo matching cost, deep learning, unsupervised training, unlabeled data.**

## INTRODUCTION

A STEREO matching algorithm consists of four main steps: data cost, aggregation cost, optimization, and post processing; the data cost is the most important step among the four. Many algorithms have been proposed to produce a robust stereo data cost under various conditions. These include the Census and Rank transforms [1], adaptive normalized cross correlation (ANCC) [2], guided image filtering [3], and local density encoding (LDE) [4]. The main objective of those data costs is to extract more accurate and robust information from the local region. However, expert knowledge is required to develop those data costs under specific conditions. For example, Census works well when the gray-level changes monotonically; however, it fails in uniform regions or with larger illumination variations. ANCC can handle neither multiple illumination conditions nor non-Lambertian reflectance objects. Thus, those data costs cannot function well under conditions that vary from those for which they were designed. A combination of unknown factors, such as rain, snow, and sun, is significantly challenging to the development of a robust stereo data cost in the real world, which raises an interesting question. Is it possible to develop a robust data cost model that does not require much expert knowledge under a specific condition? Recently, deep learning performed optimally in various applications, such as natural language processing, texture classification, and object recognition [5]–[7]. The performance of existing applications has been significantly increased, by 30% or more, through the use of the deep learning approach. Therefore, this paper investigates a deep learning strategy to improve the performance of existing stereo matching algorithms. The main contributions of this paper are as follows.

1) This research is the first to apply a deep learning strategy-based unsupervised approach to develop a robust transformation function to compute stereo matching costs (it is different from the work in [8], which used supervised learning to develop a new data cost function).
2) We investigate how a deep learning strategy can be applied to calculate stereo data costs.
3) The proposed transformation function outperforms existing state-of-the-art data costs using the KITTI benchmark [9] and the Middlebury [10] and HCI datasets [11].
4) The proposed approach can improve the performance of existing stereo data costs without requiring their structures to be modified.

The remainder of this paper is organized as follows. Section II provides a brief summary of existing stereo matching algorithms and their limitations. The proposed cost function for stereo matching is presented in Section III. Our experimental results are presented in Section IV. Finally, the conclusion and proposed future work are presented in Section V.

## II. RELATED WORK

Stereo vision is an important factor used to detect and estimate the distance to the preceding vehicle in driving assistance systems [12]. Therefore, many stereo matching algorithms have been proposed to produce an accurate disparity map [13]–[19]. Without loss of generality, stereo matching algorithms can be divided into three categories: local, semi-global/non-local, and global algorithms. Local algorithms often provide fast processing time suitable for real-time systems. Global algorithms are usually more accurate than local algorithms, but they require

AI VIETNAM
All-in-One Course

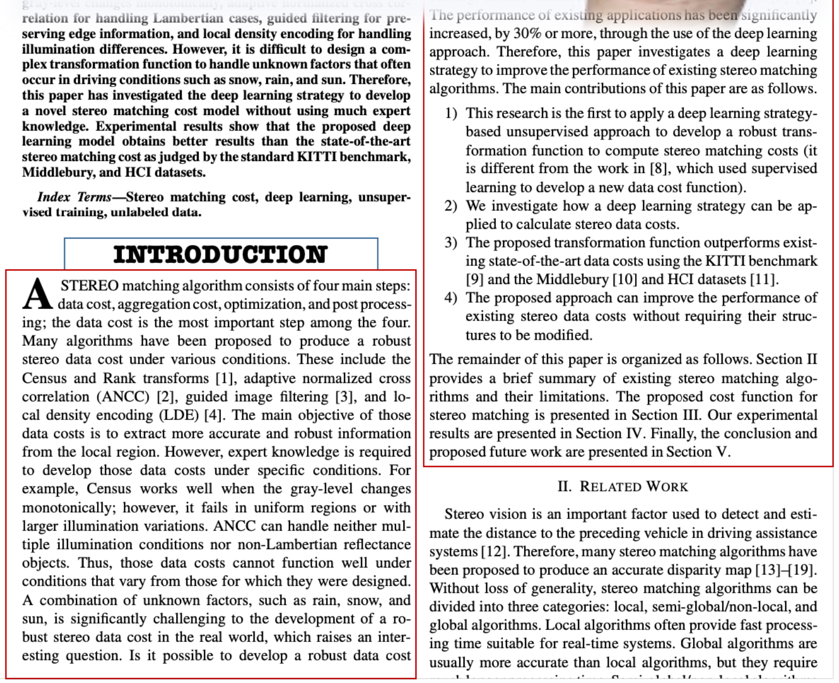
# How to Write an Introduction

1. BACKGROUND OF RESEARCH TOPICS/QUESTIONS

2. STRENGTH AND WEAKNESS OF EXISTING METHODS

3. RESULTS OF YOUR PROPOSED METHOD vs. SOTA METHODS

4. MAIN CONTRIBUTIONS OF YOUR PROPOSED METHOD

**relation for handling Lambertian cases, guided filtering for preserving edge information, and local density encoding for handling illumination differences. However, it is difficult to design a complex transformation function to handle unknown factors that often occur in driving conditions such as snow, rain, and sun. Therefore, this paper has investigated the deep learning strategy to develop a novel stereo matching cost model without using much expert knowledge. Experimental results show that the proposed deep learning model obtains better results than the state-of-the-art stereo matching cost as judged by the standard KITTI benchmark, Middlebury, and HCI datasets.**

***Index Terms*—Stereo matching cost, deep learning, unsupervised training, unlabeled data.**

INTRODUCTION

A STEREO matching algorithm consists of four main steps: data cost, aggregation cost, optimization, and post processing; the data cost is the most important step among the four. Many algorithms have been proposed to produce a robust stereo data cost under various conditions. These include the Census and Rank transforms [1], adaptive normalized cross correlation (ANCC) [2], guided image filtering [3], and local density encoding (LDE) [4]. The main objective of those data costs is to extract more accurate and robust information from the local region. However, expert knowledge is required to develop those data costs under specific conditions. For example, Census works well when the gray-level changes monotonically; however, it fails in uniform regions or with larger illumination variations. ANCC can handle neither multiple illumination conditions nor non-Lambertian reflectance objects. Thus, those data costs cannot function well under conditions that vary from those for which they were designed. A combination of unknown factors, such as rain, snow, and sun, is significantly challenging to the development of a robust stereo data cost in the real world, which raises an interesting question. Is it possible to develop a robust data cost

The performance of existing applications has been significantly increased, by 30% or more, through the use of the deep learning approach. Therefore, this paper investigates a deep learning strategy to improve the performance of existing stereo matching algorithms. The main contributions of this paper are as follows.

1) This research is the first to apply a deep learning strategy-based unsupervised approach to develop a robust transformation function to compute stereo matching costs (it is different from the work in [8], which used supervised learning to develop a new data cost function).
2) We investigate how a deep learning strategy can be applied to calculate stereo data costs.
3) The proposed transformation function outperforms existing state-of-the-art data costs using the KITTI benchmark [9] and the Middlebury [10] and HCI datasets [11].
4) The proposed approach can improve the performance of existing stereo data costs without requiring their structures to be modified.

The remainder of this paper is organized as follows. Section II provides a brief summary of existing stereo matching algorithms and their limitations. The proposed cost function for stereo matching is presented in Section III. Our experimental results are presented in Section IV. Finally, the conclusion and proposed future work are presented in Section V.

II. RELATED WORK

Stereo vision is an important factor used to detect and estimate the distance to the preceding vehicle in driving assistance systems [12]. Therefore, many stereo matching algorithms have been proposed to produce an accurate disparity map [13]–[19]. Without loss of generality, stereo matching algorithms can be divided into three categories: local, semi-global/non-local, and global algorithms. Local algorithms often provide fast processing time suitable for real-time systems. Global algorithms are usually more accurate than local algorithms, but they require

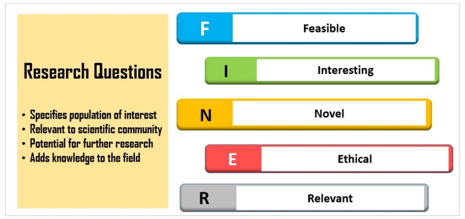
# How to Write an Introduction

## 1. BACKGROUND OF RESEARCH TOPICS/QUESTIONS

**Springer format:**
https://www.springer.com/gp/authors-editors/conference-proceedings/editors/word-template/19338734

**IEEE Format:**
https://www.ieee.org/conferences/publishing/templates.html

- **Assigment 1:**
  - Requirements: Find and read 10 papers
  - Create a summarized table
  - Reseach content: your interested research
  - Deadline: **10/03/2023**
  - Submission to email: **tamnguyen100694@gmail.com**



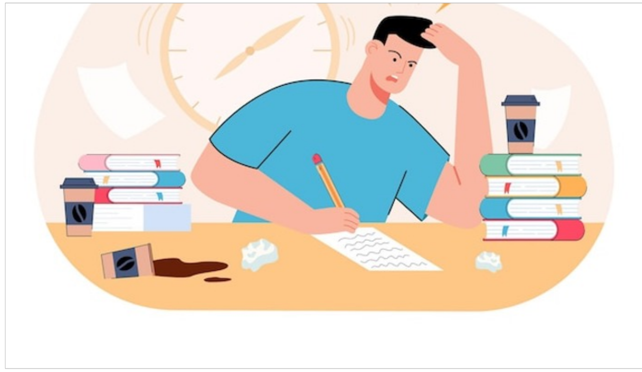

- **Assigment 2:**
  - Requirements: **Describe three to five research questions**
  - Reseach content: your interested research
  - Deadline: **18/03/2023**
  - **Submission to email: aivnresearch@gmail.com**

# How to Write an Introduction

## 2. STRENGTH AND WEAKNESS OF EXISTING METHODS

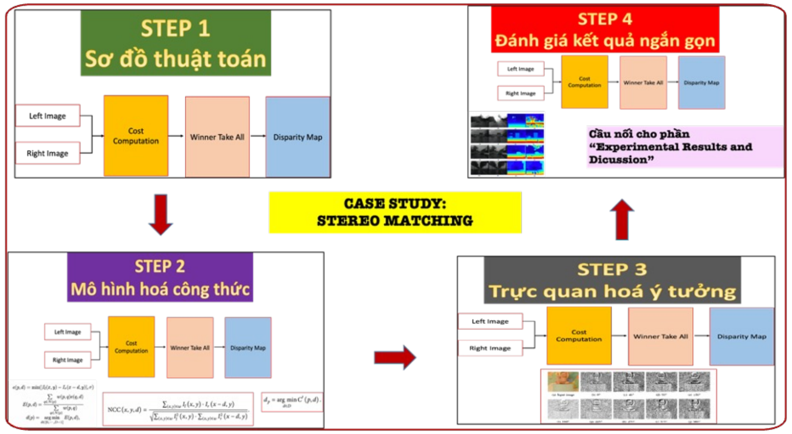
**Springer format:**
https://www.springer.com/gp/authors-editors/conference-proceedings/editors/word-template/19338734

- **Assignment 3:**
  - Requirements: **Write Related Work and References Section**
  - Research content: your interested research
  - Deadline: **25/03/2023**
  - Submission to email: **aivnresearch@gmail.com**



# How to Write an Introduction

## 3. MAIN CONTRIBUTIONS OF YOUR PROPOSED METHOD

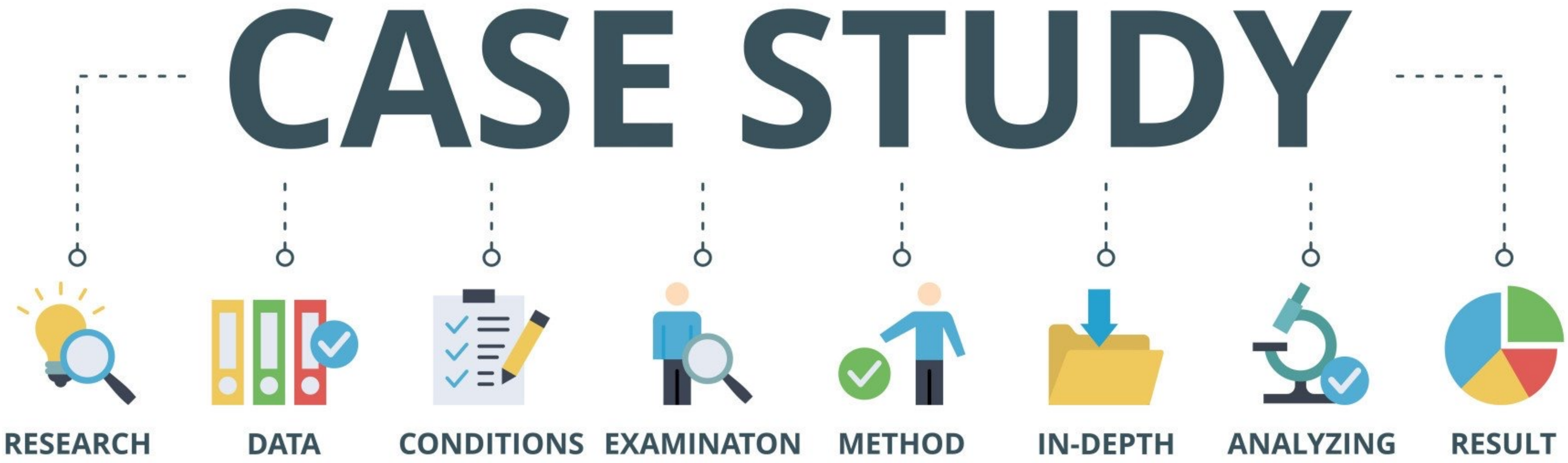



- **Assignment 4:**
  - Requirements: **Write The Proposed Method Section**
  - Research content: your interested research
  - Deadline: **01/04/2023**
  - Submission to email: **aivnresearch@gmail.com**

# How to Write an Introduction

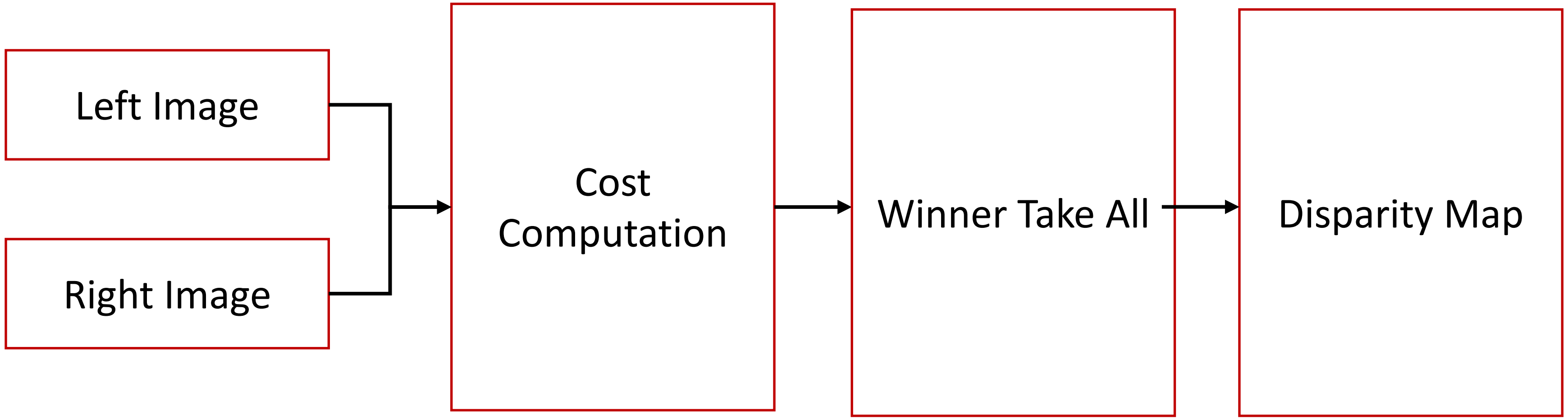
## 4. RESULTS OF YOUR PROPOSED METHOD Vs. SOTA METHODS

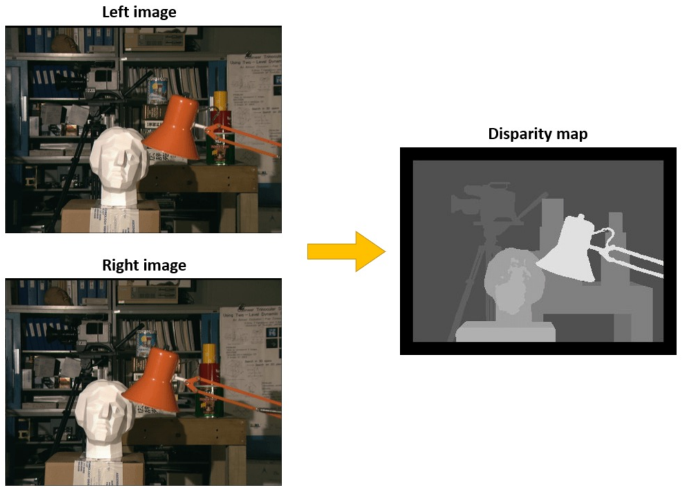
- **Assignment 5:**
  - Requirements: **Write The Experimental Results Section**
  - Research content: your interested research
  - Deadline: **08/04/2023**
  - Submission to email: **aivnresearch@gmail.com**


CASE STUDY
RESEARCH
DATA
CONDITIONS
EXAMINATON
METHOD
IN-DEPTH
ANALYZING
RESULT

# How to Write an Introduction

## INTRODUCTION FOR STEREO MATCHING TOPIC



# How to Write an Introduction

## STEP 1
## BACKGROUND OF RESEARCH TOPICS/QUESTIONS

Stereo vision plays an important role in many applications, such as 3-D tracking [1], 3-D driving environments [2], 3-D scanning [3], and 2-D and 3-D change detection [4]. Stereo information is effective for detecting vehicles and estimating distances to vehicles in driving conditions, as described in a comprehensive report [5]. However, stereo vision-based driving assistance is challenging under realistic conditions due to difficulties such as occlusion, textureless regions, illumination artifacts, flying snow, rain flare, rain blur, sun flare, or illumination differences, as described in the Heidelberg Collaboratory for image processing (HCI) data set [6].

# How to Write an Introduction

## STEP 2
## STRENGTH AND WEAKNESS OF EXISTING METHODS

STEREO matching algorithm consists of four main steps: data cost, aggregation cost, optimization, and post processing; the data cost is the most important step among the four. Many algorithms have been proposed to produce a robust stereo data cost under various conditions. These include the Census and Rank transforms [1], adaptive normalized cross correlation (ANCC) [2], guided image filtering [3], and local density encoding (LDE) [4]. The main objective of those data costs is to extract more accurate and robust information from the local region. However, expert knowledge is required to develop those data costs under specific conditions. For example, Census works well when the gray-level changes monotonically; however, it fails in uniform regions or with larger illumination variations. ANCC can handle neither multiple illumination conditions nor non-Lambertian reflectance objects. Thus, those data costs cannot function well under conditions that vary from those for which they were designed.

# How to Write an Introduction

## STEP 3
## RESULTS OF YOUR PROPOSED METHOD Vs. SOTA METHODS

- Many stereo methods are based on the assumption that the colors or intensities of the left and right images have similar values and perform very well-under controlled conditions [10]–[15]; however, their performances degrade significantly in the presence of photometric or radiometric variations, including illumination direction, illuminant color, brightness differences, and image device changes. Fig. 2 shows the performance of the new method proposed in this paper with two conventional global and semiglobal stereo methods under rain flare conditions: semiglobal matching (SGM) [8] and belief propagation (BP) [10] using the sum of absolute

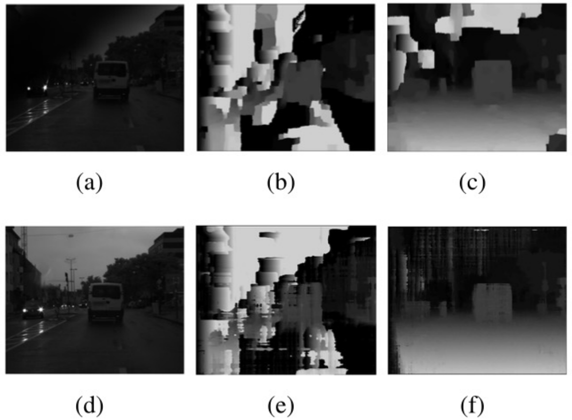



Fig. 2. Experimental results of the proposed method with existing stereo methods under brightness difference conditions. (a) and (d) Left and right images under rain flare condition in the HCI data set, respectively. (b) and (e) Disparity results of (BP+SAD) and (SGM+SAD), respectively. (c) and (f) Disparity results of the proposed LDE method with BP (BP+LDE) and the proposed LDE method with SGM (SGM+LDE), respectively.

# How to Write an Introduction

## STEP 4
## MAIN CONTRIBUTIONS OF YOUR PROPOSED METHOD

Recently, deep learning performed optimally in various applications, such as natural language processing, texture classification, and object recognition [5]–[7]. The performance of existing applications has been significantly increased, by 30% or more, through the use of the deep learning approach. Therefore, this paper investigates a deep learning strategy to improve the performance of existing stereo matching algorithms. The main contributions of this paper are as follows.

1) This research is the first to apply a deep learning strategy- based unsupervised approach to develop a robust trans- formation function to compute stereo matching costs (it is different from the work in [8], which used supervised learning to develop a new data cost function).

2) We investigate how a deep learning strategy can be ap- plied to calculate stereo data costs.

3) The proposed transformation function outperforms existing state-of-the-art data costs using the KITTI benchmark [9] and the Middlebury [10] and HCI datasets [11].

4) The proposed approach can improve the performance of existing stereo data costs without requiring their structures to be modified.

# How to Write an Introduction

## ENDING
## MAIN CONTRIBUTIONS OF YOUR PROPOSED METHOD

The remainder of this paper is organized as follows. Section II provides a brief summary of existing stereo matching algorithms and their limitations. The proposed cost function for stereo matching is presented in Section III. Our experimental results are presented in Section IV. Finally, the conclusion and proposed future work are presented in Section V

# How Do Reviewers Expect

## Clarity of Purpose

Reviewers look for a clear statement of the research problem, question, or hypothesis. The introduction should succinctly convey the purpose and significance of the study.

## Background Information

A good introduction provides sufficient background information to help readers understand the context of the research. Reviewers assess whether the background is relevant and well-balanced, offering enough information without overwhelming the reader.

## Literature Review

Reviewers examine the integration of the literature review into the introduction. They look for a critical review of existing research, demonstrating the author's understanding of the current state of the field.

## Research Gap Identification

The introduction should identify a gap in the existing literature that the current study aims to address. Reviewers assess how well the author justifies the need for their research based on identified gaps.

## Significance of the Study

Reviewers assess how well the introduction communicates the broader significance of the research. It should articulate why the study is important and how it contributes to the field.

## Logical Flow

The introduction should have a logical flow, guiding the reader from the general context to the specific research question. Reviewers check for coherence and smooth transitions between different components of the introduction.

# Reviewers' Comments

Reviewer Comment: "The introduction is somewhat verbose. Consider condensing the background information and focusing on key points to enhance clarity and conciseness."

Reviewer Comment: "The research question and objectives are not clearly articulated. Please explicitly state your research aim and the specific objectives guiding your study."

Reviewer Comment: "The introduction lacks sufficient contextualization of the research problem. Provide more background information to help readers understand the significance of the study."

Reviewer Comment: "Integrate the literature review more seamlessly into the introduction to establish a stronger foundation for your research. Connect the existing knowledge to the gap your study aims to address."

Reviewer Comment: "There is some redundancy with the related work section. Ensure that the introduction and related work complement each other without repeating information."

# Outline

- How to Write Related Work Section
- How to Write Introduction Section
- Assignment

Assignment 3:

- Requirements: Write The Related Work and Introduction Sections
- Research content: your interested research
- Deadline: 03/02/2023
- Submission to email: aivnresearch@gmail.com

# Research Paper Title that You Plans to Submit to International Conference or Journal *

*How to write a scientific research paper: Assignment 2 - Write the Experimental results Sections

$1^{st}$ Given Name Surname
*dept. name of organization (of Aff.)*
*name of organization (of Aff.)*
City, Country
email address or ORCID

$2^{nd}$ Given Name Surname
*dept. name of organization (of Aff.)*
*name of organization (of Aff.)*
City, Country
email address or ORCID

$3^{rd}$ Given Name Surname
*dept. name of organization (of Aff.)*
*name of organization (of Aff.)*
City, Country
email address or ORCID

$4^{th}$ Given Name Surname
*dept. name of organization (of Aff.)*
*name of organization (of Aff.)*
City, Country
email address or ORCID

$5^{th}$ Given Name Surname
*dept. name of organization (of Aff.)*
*name of organization (of Aff.)*
City, Country
email address or ORCID

$6^{th}$ Given Name Surname
*dept. name of organization (of Aff.)*
*name of organization (of Aff.)*
City, Country
email address or ORCID

***Abstract*—[Will be submitted in next assignments]**
***Index Terms*—component, formatting, style, styling, insert**

## I. Introduction

[Will be submitted in next assignments]

## II. Related Work

[Will be submitted in next assignments]

## III. The Proposed Method

[Finished]

## IV. Experimental Results

Please complete this section and send it to AIVN Research Team in order to receive feedback and suggestion in improving the quality of the research.

### *A. Research Questions*

Research Question 1? Research Question 2? Research Question 3?

### *B. System configuration, Dataset and Evaluation Metrics*

### *C. Results and Discussion*

*1) Research Question 1: Answer and Discussion:*
*2) Research Question 2: Answer and Discussion:*
*3) Research Question 3: Answer and Discussion:*

### *D. Limitations*

## V. Conclusions

[Will be submitted in next assignments]

How to write a scientific research paper: Assignment 2 - Write the Experimental results Sections

## Acknowledgment

[Will be submitted in next assignments]

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks ...". Instead, try "R. B. G. thanks...". Put sponsor acknowledgments in the unnumbered footnote on the first page.

## References

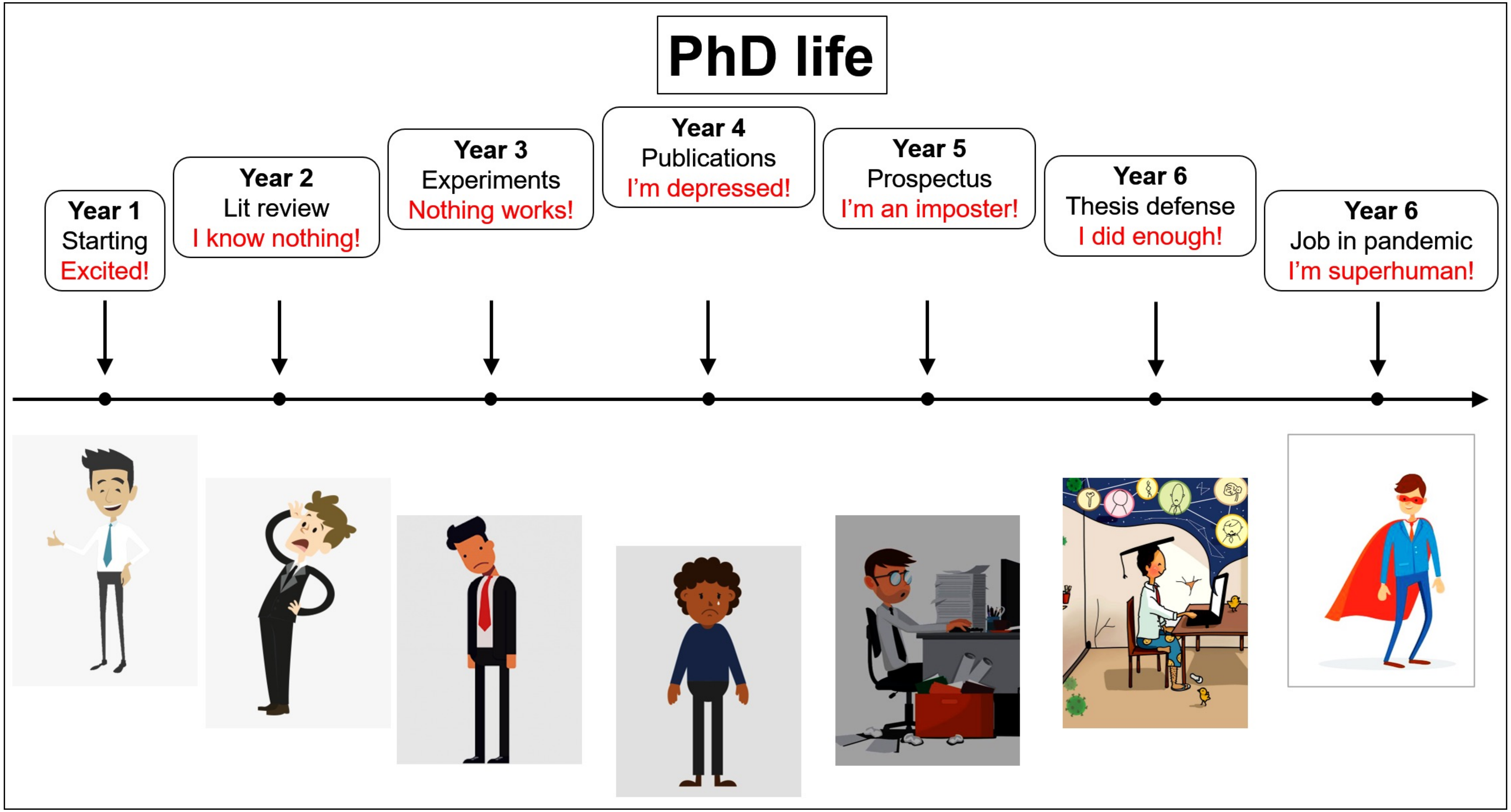
[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[4] K. Elissa, "Title of paper if known," unpublished.
[5] R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

*********************************************

## VI. Guide to Write

The IEEEtran class file is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an

# PhD Life


PhD life
Year 1
Starting
Excited!
Year 2
Lit review
I know nothing!
Year 3
Experiments
Nothing works!
Year 4
Publications
I'm depressed!
Year 5
Prospectus
I'm an imposter!
Year 6
Thesis defense
I did enough!
Year 6
Job in pandemic
I'm superhuman!


Thank you


Thank you