עבודת סיקור מאמר - מבוא לעיבוד ספרתי של תמונות

שם המגיש: תום לב-רון.

שם מאמר: Scale-recurrent Network for Deep Image Deblurring.

# מבוא

טשטוש תמונה הוא אחד מן תוצרי הלוואי הנפוצים כאשר מצלמים אובייקט. תנועות חדות של האובייקט או המצלמה יכולים להוריד את איכות התמונה ולגרום לטשטוש. בנוסף גורמים נוספים כמו עומק תמונה משתנה שלא נרכש בעומק השדה הרצוי וכדומה יכולים להערים קשיים נוספים. בעבר מחקרים נפוצים היו בעיקר על טשטושי של תזוזות המצלמה, כיום מחקרים מתמקדים בסוגים שונים של טשטושים הנובעים מטשטוש לא אחיד. שיטות תיקון מקובלות ונפוצות מניחות ידע מוקדם על פונקציית המריחה או ניסיון לשערך אותו, שיטות אלו מגבילות מאוד את מאפייני שערוך הטשטוש. בעיה נוספת היא שלעיתים החישובים האלו לא מוציאים תוצאות טובות על מידע שאינו מהמעבדה, כלומר מידע "בעולם האמיתי". הסיבה לזה היא שבעולם האמיתי הטשטוש לעיתים יותר מסובך ומגוון ולעיתים מושפע מחישובים של חומרת המצלמה.

ההתמקדות הנפוצה כיום במחקרים היא בשחזור תמונות ללא ידיעה של פונקציית המריחה, שיטה זו מכונה "תיקון טשטוש עיוור". שיטות למידה שונות הוצעו אף הם לתיקון טשטושים בתמונות על ידי שימוש במידע חיצוני בדומה למאמר זה אך לא בהכרח על ידי רשת. המאמר מציע שיטה לתיקון טשטושים הנובעים ממספר גורמים שונים כמו תנועה של האובייקט, מיקוד לא נכון של העדשה, רעידה של המצלמה וכדומה. כותבי המאמר מציינים לשבח עבודה אחרת בתחום עליה דיברנו בהרצאות שלנו וטוענים לשיפור משמעותי בה [‎7]. הבעיה שכותבי המאמר הנוכחי טוענים שיש במחקרים אחרים היא שהפתרון והפרמטרים של כל משקול קרנל תנועה הם בדרך כלל זהים. הכותבים מציעים שימוש בפרמטרים של הרשת עבור כמה משקולים שונים של טשטוש. **עצרתי במרקור האחרון של המאמר**

# מטרה

המטרה היא לשחזר את התמונה כך שתתקבל תמונה חדשה וחדה. הכיוון של המחקר הוא תיקון טשטוש בתמונה יחידה בצורה כזו שמשחזרים את התמונה החדה בעזרת מעבר בין רזולוציות שונות מהקטנות לגדולות. המאמר בעצם מציע שיטה לשחזור תמונות מטושטשות על ידי שימוש ברשת נוירונים מלאכותית בסקלות שונות של התמונה.

# שיטות

הבעיה איתה מנסים החוקרים להתמודד היא בעיה מוכרת אשר יכולה להיפתר במגוון שיטות. ישנם השיטות "הקלאסיות", אשר משתמשות בפילטרים ושערוך פונקציית הטשטוש. ישנם שיטות בהם משערכים את פונקציית הטשטוש על ידי למידה ולאחר מכן מסננים באמצעותה בעזרת פילטרים מוכרים. בשנים האחרונות ישנו דגש גדול על שימוש בשיטות למידה להסרת טשטוש מתמונות בצורה עיוורת (ללא ידע מוקדם על התמונה).

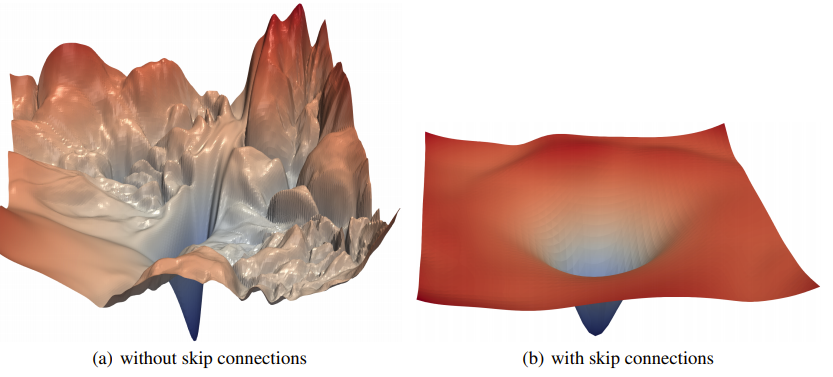
## שיטות קיימות

## השיטה המוצעת במאמר

השיטה המוצעת היא שימוש ברשת נוירונים אשר מקבלת תמונות בגודל מסוים (במקרה הזה רצף של תמונות), מקטינה אותם ומחלקת אותם ל3 קני מידה שונים ("סקלות") של גודל התמונה, שיטת הפירמידה היא כלי מוכר בשיטות ומשימות שחזור מטשטוש או יצירת תמונות [[‎7,‎4. כלומר יש כאן מעין תהליך פירמידה בשינוי גודל התמונה תוך כדי ניבוי של תמונה חדה מתמונה מטושטשת מכל קנה מידה כתת משימה. בין כל קנה מידה לקנה מידה ישנו חיבור של הפלט מהקנה מידה הקודם אל הקלט של הקנה מידה החדש (הפלט מהקנה מידה קודם עובר הגדלה כך שיתאים לרזולוציה של קנה המידה החדש בעזרת אינטרפולציה בי-לינארית).

החוקרים מציעים שימוש בשיטת הפירמידה בשילוב עם רשת מסוג RNN (Recurrent Neural Network), שיטות קודמות נעשו בעזרת CNN (Convolutional Neural Networks) ברובם. בנוסף ישנו שימוש בארכיטקטורת מקודד-מפענח (Encoder-Decoder) שידועה בשימושה בבעיות הפוכות [‎8]. המקודד יוצר מפות מאפיינים של המידע המוזן אליו והמפענח בונה מידע חדש מן מפות המאפיינים שהוזנו אליו מהמקודד. במקרה של המאמר הזה, המקודד בונה מפות מאפיינים של הטשטוש והמפענח ינסה לבנות תמונה מתאימה שאינה מטושטשת בעזרת מפות המאפיינים. הRNN בא פה לידי ביטוי בכך שישנו בלוק LSTM (Long-Short-Term-Memory) בין המקודד והמפענח בנוסף לשימוש בפלט מהקנה המידה הקודם כקלט נוסף עבור הקנה מידה הבא, הבלוק בעצם יוצר סוג של שיתוף פרמטרים בין קני המידה השונים. דבר נוסף שמתרחש בין כל קנה מידה לקנה מידה הוא שיתוף פרמטרים, זה לא מתרחש בדרך כלל בCNN ויכול לעזור לאימון מהיר ויעיל יותר של הפרמטרים. היתרון שנוצר מכך הוא שניתן לחלוק את הפרמטרים של כל "הסקלות" ולא לחשב פרמטרים שונים עבור כל סקלה וכך לחסוך זיכרון, זמן אימון וזמן ניבוי.

המקודד-מפענח שהחוקרים משתמשים בו עושה שימוש בResBlocks אשר עוזרים מאוד להימנע מדיפוזיה של גרדיאנטים ולהוביל להתכנסות יותר מהירה וטובה [‎5,‎6]. כך נוכל בעצם להעמיק את הרשת ולהשתמש בעוד שכבות. בנוסף ישנם "קשרי קפיצה" בין המקודד והמפענח, אלו קשרי קפיצה "טווח ארוך" אשר בעצם מחברים מפות מאפיינים אל מפות מאפיינים של המפענח וכך בעצם להעביר מאפיינים משכבות מוקדמות למאוחרות.



איור 1: משטח פונקציית ההפסד, צד שמאל ללא קשרי קפיצה וצד ימין עם קשרי קפיצה.

כמו שנאמר, הקלט של הרשת הוא רצף של תמונות מטושטשות והמטרה שלה היא לנבא את התמונה "הנקייה" של כל תמונה כרצף תמונות. הקלט נחתך בצורה רנדומלית ומוקטן לרבע מהגודל המקורי. כלומר ניתן לרשום זאת כמשוואה:

כאשר I זו התמונה המשוחזרת משוערכת, h אלו וקטורים פנימיים ברשת מתוך בלוק הLSTM, B זו התמונה המטושטשת, "תטא" אלו פרמטרי האימון ברשת. הווקטורים הפנימיים הם מאפיינים מסקלות קודמות. פונקציית האקטיבציה של הרשת היא ReLU. ה"פותר" של הרשת הוא Adam הם פונקציית הפסד L2: . *קצב הלימוד*  *דועך אקספוננציאלית.*

# דיון

לדעתי בעיה שנוצרת מכך שיש שיתוף פרמטרים בין הסקלות היא שיש מאפיינים ספציפיים לכך סקלה וייתכן ונאבד אותם. אולי עדיף לבצע בחירה של פרמטרים, קצת יותר קשה אבל ראיתי שכבר עושים את זה.

# סיכום

# נספחים

## CNN

רשת CNN בעצם לומדת לבנות מפות מאפיינים מתוך הקלט שמגיע במקרה שלנו מתמונות, שכבת קונבולוציה תחשב את התוצאה של נוירונים המחוברים לאזורים מקומיים בתמונה, כל אחד מהם יחשב מכפלה סקלרית בין המשקולות ובין האזור שאליו הם מחוברים בתמונה. אם נניח נבחר עומק של 12 פילטרים נקבל מבנה תלת מימדי של 12 מפות מאפיינים על גודל התמונה שנניח 256X256X12. כל פילטר בעצם עובר על התמונה ומבצע קונבולוציה על אזור בה ובונה מפת דו-מימד שהיא הפילטר. כלומר רשת CNN משנות את התמונה המקורית שכבה-שכבה כך שנלמדים מאפיינים וככל שהרשת עמוקה יותר נלמד יותר מאפיינים.

## Encoder-Decoder

## RNN

בCNN הניבוי של עצם אחד לא בהכרח מתקשר לניבוי של עצם אחר, כלומר אם המידע נכנס בצורה רציפה אין זיכרון במערכת ללמוד ממה שהיה בעבר. RNN מבצע את אותה המשימה עבור כל אלמנט בסדרה שנכנסה כאשר הפלט תלוי בחישובים קודמים. לRNN יש משקולות משותפות לאורך התהליך ולכן יכול "לזכור" מידע שנצבר. RNN נוטה לגרום לגרדיאנטים להתפוצץ או להיעלם יותר מהר מCNN לכן נעשה שימוש בעוד תוספות בדרך כלל וSKIP. הסיבה לכך היא שבRNN הפרמטרים משותפים לעומת CNN.

## LSTM

נועד לפתור את הבעיה שנוצרת בRNN כאשר הגרדיאנטים גדלים מאוד או קטנים מאוד ולכם נעלמים או מתפוצצים וחיזוי משפט למשל לומד יותר מסוף המשפט מאשר תחילת משפט. שיטה נוספת שמשתמשים בה כתוספת היא קיצוץ גרדיאנטים. יש 3 שערים, שער שכחה, שער קלט ושער פלט. שער השכחה מחליט כמה מהעבר אתה רוצה לזכור על ידי פונקציית סיגמואיד ועל פי המצב של התא הקודם והקלט הנוכחי מוציא מספר בים 0 ל1 עבור כל מספר במצב התא הקודם. שער הקלט מחליט כמה מהיחידה הנוכחית יתווסף למצב הנוכחי בתא. על ידי סיגמואיד וגם טנגס היפרבולי נקבל משקול בין -1 ל1 כמה יש חשיבות לקלט עבור התא הנוכחי ולהעביר הלאה. שער פלט מחליט איזה חלק במצב התא הנוכחי יועבר לפלט על פי חשיבות. C זה תא הזיכרון הנצבר, הוא צובר את מצב האינפורמציה הנוכחית. התא נכתב, מונגש ונערך על ידי פרמטרים מוגדרים ושערים.

כעת, עבור תא מסוג קונבולוציה, ניתן להתייחס גם לפרמטרים המרחביים בכל שכל הקלטים, המצבים, שערים ומצבים חבויים הם בצורת מידע ספקטרלי תלת ממדי כך שניתן להתייחס למידע דו ממדי. הטנגגנטH דואג לפיזור הגרידאנטים כדי למנוע פיצוץ או איפוס. h זוהי מטריצה של ערכים חבויים שהם בעצם הפלט של תאים קודמים. b זה הBIAS, Wx זה מטריצת משקולות.

## Resblock

# מקורות

1. Bahat, Yuval, Netalee Efrat, and Michal Irani. "Non-uniform blind deblurring by reblurring." Proceedings of the IEEE International Conference on Computer Vision. 2017.‏
2. Cho, Sunghyun, and Seungyong Lee. "Fast motion deblurring." ACM SIGGRAPH Asia 2009 papers. 2009. 1-8.‏
3. Cho, Sunghyun, Yasuyuki Matsushita, and Seungyong Lee. "Removing non-uniform motion blur from images." 2007 IEEE 11th International Conference on Computer Vision. IEEE, 2007.‏
4. Denton, Emily, et al. "Deep generative image models using a laplacian pyramid of adversarial networks." arXiv preprint arXiv:1506.05751 (2015).
5. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 770–778 (2016).
6. Li, H., Xu, Z., Taylor, G., Studer, C. & Goldstein, T. Visualizing the loss landscape of neural nets. Adv. Neural Inf. Process. Syst. 2018-December, 6389–6399 (2018).
7. Nah, Seungjun, Tae Hyun Kim, and Kyoung Mu Lee. "Deep multi-scale convolutional neural network for dynamic scene deblurring." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.‏
8. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9351, 234–241 (2015).
9. Sun, Jian, et al. "Learning a convolutional neural network for non-uniform motion blur removal." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.‏
10. Tao, Xin, et al. "Scale-recurrent network for deep image deblurring." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.‏