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 מתוצרי הלוואי הנפוצים כאשר מצלמים אובייקט. תנועות חדות של האובייקט ו/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. מחקרים נפוצים מתעסקים בעיקר בטשטושי 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 לא מוציאים תוצאות טובות על מידע שאינו מהמעבדה, כלומר מידע "בעולם האמיתי". הסיבה לזה היא שבעולם "האמיתי" הטשטוש לעיתים יותר מסובך, מגוון ולעיתים מושפע מחישובים של חומרת המצלמה. המאמר שאסביר מנסה להציג גישה מעט שונה וינסה להיות כללי מספיק כדי לפתור תמונות מהחיים ה"אמתיים".

# 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

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. ישנם שיטות רבות לתיקון טשטוש בתמונה, למדנו על כך בהרצאה. פילטר "ווינר" יכול לשמש לתיקון טשטוש עם צירוף ידע מוקדם או שערוך פונקציית הטשטוש במערכת. שימוש בלמידה של פונקציית הטשטוש ולאחר מכן להשתמש בפילטרים. נעשו מחקרים לפני המחקר הזה על תיקון טשטוש באמצעות רשת מסוג CNN. פילטרים הופכיים ידועים בשימוש עבור לוחות רישוי בהם הטשטוש הוא העיקר והרעש פחות גורם משפיע.

## השיטה המוצעת במאמר

המאמר מציע שיטה לתיקון טשטושים הנובעים ממספר גורמים שונים כמו תנועה של האובייקט, מיקוד לא נכון של העדשה, רעידה של המצלמה וכדומה. הבעיה שכותבי המאמר הנוכחי טוענים שיש במחקרים אחרים היא שהפתרון והפרמטרים של כל משקול קרנל תנועה הם בדרך כלל זהים.

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. 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 [[‎9,‎4. 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 [‎11]. 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. כלומר בין כל קנה מידה לקנה מידה מתקיים שיתוף פרמטרים, זה לא מתרחש בדרך כלל בCNN ויכול לעזור לאימון מהיר ויעיל יותר של הפרמטרים. היתרון שנוצר מכךהוא שניתן לחלוק את הפרמטרים של כל "הסקלות" ולא לחשב פרמטרים שונים עבור כל סקלה וכך לחסוך זיכרון, זמן אימון וזמן ניבוי.

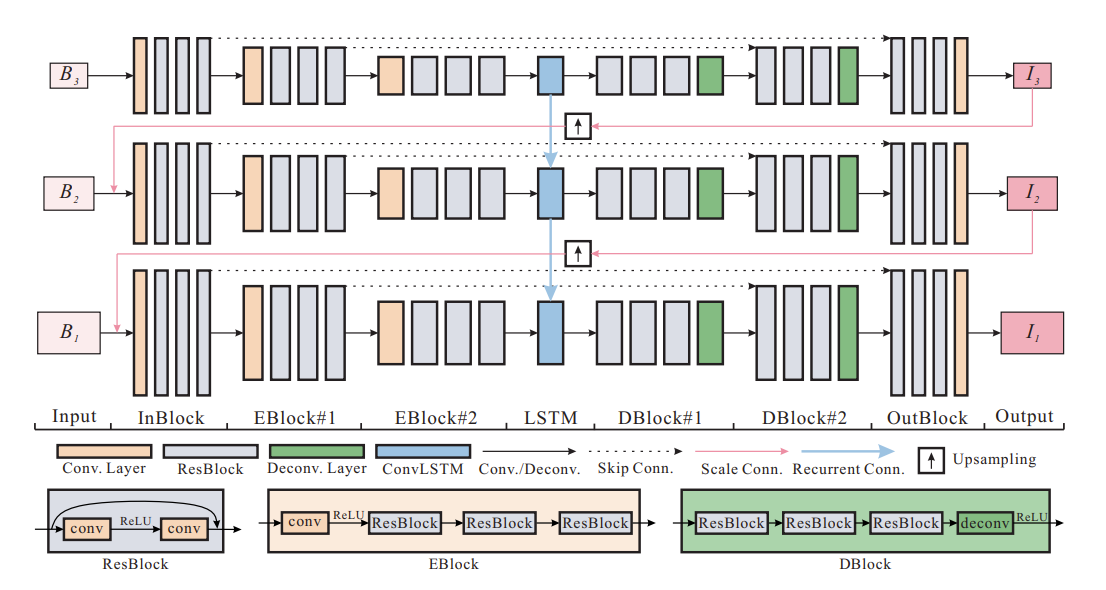
המקודד-מפענח שהחוקרים משתמשים בו עושה שימוש בResBlocks אשר עוזרים מאוד להימנע מדיפוזיה או התפוצצות של גרדיאנטים ומוביל להתכנסות מהירה וטובה יותר [‎6,‎8]. כך נוכל בעצם להעמיק את הרשת ולהשתמש בעוד שכבות. בנוסף ישנם "קשרי קפיצה" בין המקודד והמפענח, אלו קשרי קפיצה "טווח ארוך" אשר בעצם מחברים מפות מאפיינים של המקודד אל מפות מאפיינים של המפענח וכך בעצם להעביר מאפיינים משכבות מוקדמות למאוחרות. השיטה המוצעת במאמר היא ללא שימוש בשכבות מיצוע (Batch Normalization) בהתבסס על מאמרים קודמים [‎9] ובניגוד לשימוש המקורי בבלוקים [‎6]. המפענח הוא בעל שכבות דה-קונוולוציה שזוהי בעצם שכבת שחלוף של קונוולוציה שמטרתה להגדיל את הגודל המרחבי של מפות המאפיינים ולהוריד את מספר הערוצים. כל שכבות קונוולוציה ברשת הם עם קרנל בגודל 5x5 עם 2 צעדים, בResBlocks יש צעד 1 עבור הקונוולוציה. הקונוולוציה בתא הLSTM היא 3x3.

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 עם פונקציית הפסד L2 כאשר *קצב הלימוד*  *דועך אקספוננציאלית.*

הקישור בין הקני מידה השונים הוא גם מתבצע מכך שמעבירים את הקלט של סקלה מסוימת לאחרת וגם על ידי קישור ב"צוואר הבקבוק" של הרשת כפי שנאמר בLSTM. ניתן להביע את המקודד-מפענח בצורה המתמטית הבאה:

כאשר השורה הראשונה היא המקודד, השנייה שכבת הLSTM והשלישית היא המפענח.



איור 1: תיאור הרשת בצורה סכמתית, B זו התמונה המטושטשת לפי סקלות (1-3), I זו התמונה המשוערכת החדשה. ניתן לראות את "קפיצות ארוכות טווח" בין המקודד למפענח בשכבות השונות בהתאמה ואת הקישור בין הLSTM.

המידע אשר נעשה בו שימוש לאמן את הרשת מורכב מתמונות של מצלמה עם חשיפה מהירה אשר נעשה מיצוע על רצף של תמונות כדי לקבל תמונה מטושטשת. בעצם מתקבלת תמונה שמדמה חשיפה ארוכה ולכן מטושטשת, זה מדמה מצבים אמיתיים יותר טוב מאשר הכפלה במטריצת טשטוש. גודל מאגר המידע הוא 3214 תמונות כזוגות של תמונה מטושטשת ותמונה נקייה. גודל החבילה בכל איטרציה ברשת הוא 16 תמונות אשר נחתכות בצורה רנדומלית לגודל של 256 על 256 פיקסלים. כל הפרמטרים ברשת אופסו בשיטת "אקסבייר" פרט לLSTM שהוא מתחיל מאפסים.

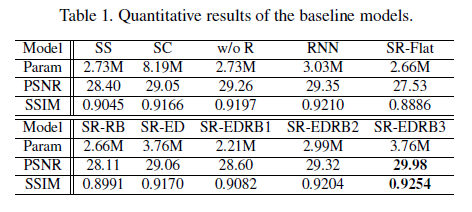
את מדידת איכות התוצאה מדדו בעזרת PSNR וSSIM:

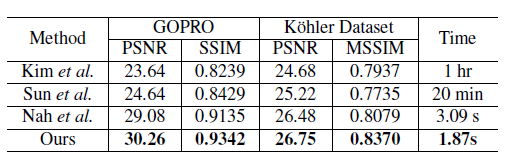
כאשר ככל שהערך של שניהם עולה ככה איכות התמונה טובה יותר. החוקרים בדקו כמה מודלים שונים: רשת אשר משתמשת בקנה מידה אחד בלבד (SS) ללא LSTM. רשת עם קנה מידה משתנה אך לא מחובר בצורה שמשתפת פרמטרים, כלומר כל קנה מידה הוא רשת בפני עצמו (SC). רשת ללא קשרי RNN באמצע הרשת (w/oR). רשת אשר משתמשת הRNN קלאסי במקום LSTM (RNN). רשת אשר לא עושה שימוש במקודד-מפענח (SR-Flat), יש במקום זה 43 שכבות קונוולוציה רציפות. רשתות אשר משתמשות במספר שונה של ResBlocks הקרויות SR-EDRB עם מספר עוקב כמספר הבלוקים. רשת אשר בבלוקים שלה אין צעדים בכלל (SR-RB). רשת מקודד-מפענח ללא ResBlocks (SR-RD).

טבלה 1 מתארת את תוצאות הניסוי על רשתות שונות. השיטה המוצעת היא SR-EDRB3. ניתן לראות שלה יש את הערכים הטובים ביותר מבחינת איכות תמונה משוחזרת.



בהשוואה לעבודות אחרות (עבודות קודמות לנוכחית) הרשת מראה שיפור גם באיכות התוצר וגם בזמן ההרצה:

טבלה 2: השוואה אל עבודות קודמות, ניתן לראות שישנו שיפור משמעותי מבחינת זמן הרצה ויחס אות-רעש.



# דיון

התוצאות מהמאמר נראות יפות, המאמר מוסבר ברור ולעניין ומצורף אליו קוד פתוח לעיון ושחזור. מאז הוצאת המאמר התפרסמו כמה וכמה מחקרים בנושא והם מסתכלים על המאמר הזה כState of the art ובסיס לבחינת השיטה שלהם [‎14,‎7]. בנוסף ישנו מאמר חדש שמשפר את השיטה שלהם על ידי טשטוש גדול יותר של התמונה [‎10].

ייתכן וישנה בעיה שנוצרת מכך שיש שיתוף פרמטרים **כולל** בין קני המידה, יש מאפיינים ספציפיים לכל סקלה וייתכן ונאבד אותם. למשל מאפיינים באזורים חדים הם דומים ואילו מאפיינים באזורים מטושטשים הם שונים בין קנה מידה אחד לאחר (ניתן לראות את זה כאשר מקטינים ומגדילים תמונה). לעיתים כשמקטינים תמונה היא נראית מעט חדה יותר וייתכן שכאשר הרשת לומדת מאפיינים בקנה המידה הנמוך היא לומדת יותר מאפיינים של חדות מאשר טשטוש. עבודה מעט חדשה יותר בנושא של כותבי המאמר מראה שישנו יתרון לשיתוף פרמטרים ברשת דומה ללא LSTM כאשר השיתוף מתבצע בצורה קצת שונה [‎5]. הכותבים מראים שיפור של 1 דציבל בPSNR כאשר בין השכבות המוציאות מאפיינים במקודדים אין שיתוף פרמטרים אך בין השכבות של הטרנספורמציות לא-לינאריות כן יש שיתוף פרמטרים. הניסוי הזה מאריך את זמן ההוצאה לפועל של הרשת ב0.3 שניות כתוצאה מהגדלת הפרמטרים אך מעלה את איכות התוצאה. הכוונה בשכבות מוציאות המאפיינים היא לשכבות קונוולוציה לפני ResBlocks ואילו הבלוקים הם הטרנספורמציות הלא-לינאריות.

# סיכום

כותבי המאמר הציגו שיטה ששוחזרה על ידי חוקרים אחרים ואף שופרה [‎10,‎5], התוצאות נעשו גם על מידע מהעולם האמיתי בו התמונות לא מטושטשות מלאכותית והראו תוצאות טובות. כותבי המאמר הסבירו את השיטה היטב ושיתפו את הקוד שלהם כך שחוקרים אחרים יוכלו לשפר ולשחזר. נעשו מספר ניסויים על מנת להראות שסידור הרשת בצורה המוצגת הוא אכן הגיוני, קצת מוזר שבניסוי שלהם הLSTM תרם הרבה ואילו בניסוי שמחקר העוקב שלהם לא נעשה בו שימוש ואף כלל לא הוזכר [‎5].

# 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. Gao, H., Tao, X., Shen, X. & Jia, J. Dynamic scene deblurring with parameter selective sharing and nested skip connections. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019-June, 3843–3851 (2019).
6. 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).
7. Kupyn, O., Martyniuk, T., Wu, J. & Wang, Z. DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. in Proceedings of the IEEE International Conference on Computer Vision 2019-October, 8877–8886 (Institute of Electrical and Electronics Engineers Inc., 2019).
8. 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).
9. 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.‏
10. Park, D., Kang, D. U. & Chun, S. Y. Blur More To Deblur Better: Multi-Blur2Deblur For Efficient Video Deblurring. (2020).
11. 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).
12. Shi, X. et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. in Advances in Neural Information Processing Systems vols 2015-Janua 802–810 (2015).
13. 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.‏
14. Tao, Xin, et al. "Scale-recurrent network for deep image deblurring." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
15. Zhang, K. et al. Deblurring by Realistic Blurring. in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2734–2743 (2020). doi:10.1109/CVPR42600.2020.00281.

# נספחים

כמה הסברים על נושאים שהרגשתי צורך להתעמק בהם כדי להבין את המאמר.

## Convolutional Neural Networks

רשת CNN בעצם לומדת לבנות מפות מאפיינים מתוך הקלט שמוזן אליה, במקרה שלנו הוא רצף של תמונות, שכבת קונוולוציה 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-ממדי של 12 מפות מאפיינים על גודל התמונה שנניח 256X256X12. כל פילטר בעצם עובר על התמונה ומבצע קונוולוציה על אזור בה ובונה מפת דו-ממד שהיא הפילטר. כלומר רשתות CNN מפעילות פעולות על התמונה המקורית ויוצרות שכבות של מפות מאפיינים וככל שהרשת עמוקה יותר נלמד יותר מאפיינים.

## Encoder-Decoder

מדובר במבנה שבדרך כלל סימטרי והוא ממפה מידע הנכנס בצורה רציפה לווקטור במקודד והמפענח בתורו לוקח את הווקטור ובונה תוצאה חדשה של רצף מידע (במקרה שלנו תמונות). המקודד יכול לקבל מידע באורך מסוים והמפענח יכול להוציא מידע באורך שונה לחלוטין.

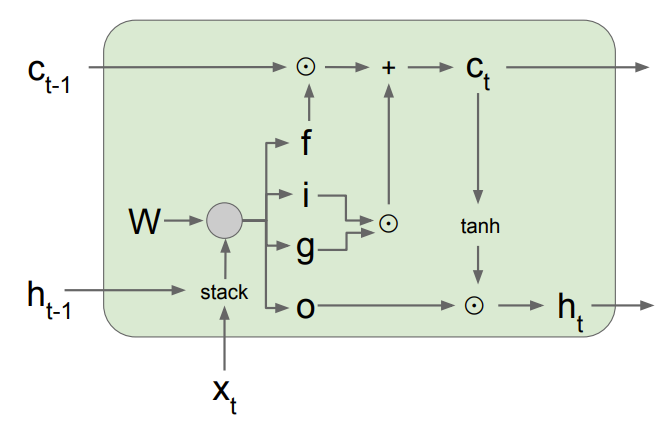
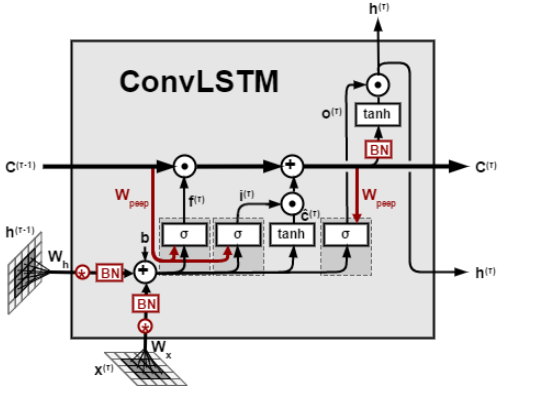
## Recurrent Neural Network

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.

## Long-Short-Term-Memory

נועד לפתור את הבעיה שנוצרת בRNN כאשר גרדיאנטים 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 מה"עבר" אתה רוצה לזכור על ידי פונקציית סיגמואיד ועל פי המצב של התא הקודם והקלט הנוכחי מוציא ערך החלטה בין 0 ל1 עבור תא הקודם. שער הקלט מחליט כמה מהיחידה הנוכחית יתווסף למצב הנוכחי בתא. על ידי סיגמואיד וגם טנגנס היפרבולי נקבל משקול בין -1 ל1 שקובע כמה יש חשיבות לקלט עבור התא הנוכחי להעביר 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 [‎12].

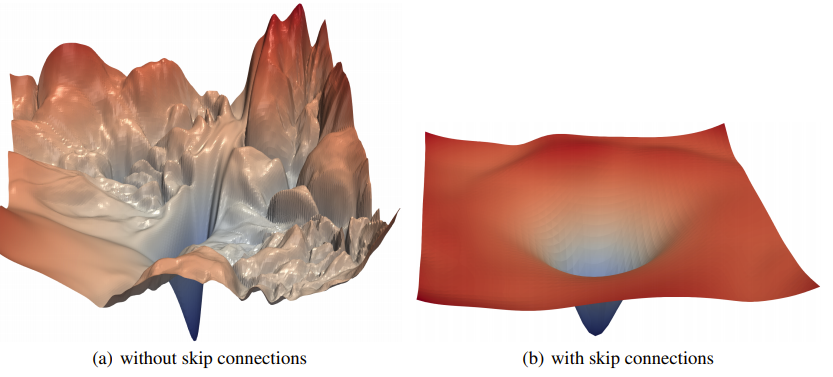
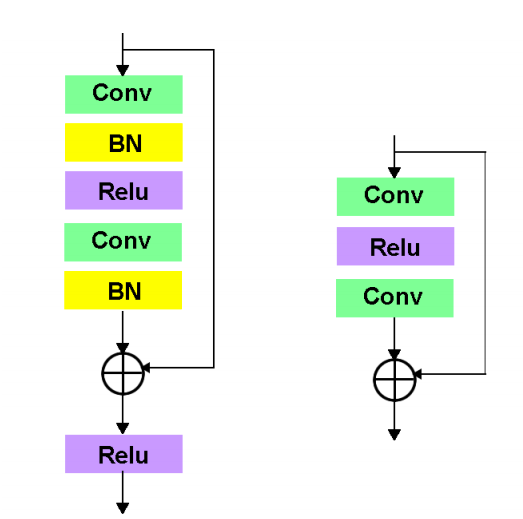
כעת, עבור תא מסוג קונבולוציה, ניתן להתייחס גם לפרמטרים המרחביים בכל שכל קלט, מצב, שער ומצב חבוי הם בצורת מידע ספקטרלי תלת ממדי כך שניתן להתייחס למידע דו ממדי. הטנגנס היפרבולי דואג לפיזור גרדיאנטים כדי למנוע פיצוץ או איפוס. h זוהי מטריצה של ערכים חבויים שהם בעצם הפלט של תאים קודמים. b זה הBIAS, Wx זה מטריצת משקולות.

איור 2: בצד ימין ניתן לראות בלוק רגיל ואילו בצד שמאל נראה בלוק קונוולוציה.

## ResBlocks

אלו בעצם בלוקים שבנויים כך שיש מספר שכבות קונוולוציה שביניהם יש פונקציית אקטיבציה ומיצוע כלשהו והעיקר כאן הוא החיבור בקשר "טווח קצר/מדלג" בין הקלט שנכנס לבלוק והפלט של הבלוק. כלומר הקלט ללא השינוי מתווסף לתוצאה של הבלוק. היתרון שמתקבל הוא יציבות והתכנסות מהירה יותר של הפתרון שמחפשים.

איור 3: משטח פונקציית ההפסד בשני האיורים הימניים, צד ימין עם קשרי קפיצה וצד שמאל ללא קשרי קפיצה. 2 האיורים השמאליים הם דוגמה לאיך הבלוק בנוי כאשר הימיני מבניהם הוא האחד שחוקרי המאמר עשו בו שימוש.