

# 3D Dense Reconstruction

Lab: CRML - The Andrew and Erna Viterbi Faculty of  
Electrical and Computer Engineering

Students: Ran Elbaz, Rinat Chanjar

Instructor: David Weinstein



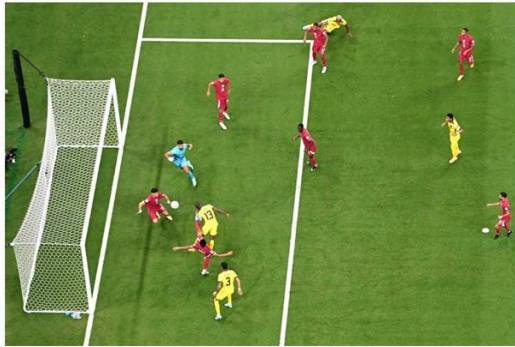
# תוכן עניינים

3.....	מבוא
4.....	מוטיבציה
Representing Scenes as Neural Radiance Fields for View	סקירת המאמר
6-7 .....	Synthesis
8-9 .....	ניסיונות לשיפור התוצאות
10.....	מסקנות ורעיונות להמשך
11.....	סיכום הפרויקט ורפלקציה
12-16 .....	נספחים

# 1. מבוא :

שחזור תלת מימד של סצנות מהעולם האמיתי הוא תחום שצובר תאוצה בשנים האחרונות. לשם קנה מידה, מספר התוצאות שנקבל ב- Google Scholar עבור חיפוש של המונח "D reconstruction3" ייתן כ-17 אלף תוצאות ב-2021 לעומת כ-55 אלף תוצאות ב-2022. המטרה בשחזור תלת מימד היא ליצור תמונות חדשות של אובייקט או סצנה מסוימים מזוויות חדשות שאינן ניתנו למערכת. כמו כן, אלמנט נוסף הוא יצירת עומק לתמונה. בדומה למגמה בהרבה תחומים, הפתרון בו מתמקדים מאמרים רבים ניתן על ידי שיטות של מערכות לומדות.

## 2. מוטיבציה:



In the third minute of play, Enner Valencia thought he had scored the first goal of the World Cup  
MANAN VATSYAYANA / AFP



The image of the 3D reconstruction of the action, as it was shown in the stadium and on television  
beIN Sports / FIFA

לשחזור מודלים תלת מימד ישנם שימושים רבים בתעשייה. כבר במונדיאל 2022 הוכנסה טכנולוגיה חדשה בה תוך דקות ספורות ניתן לראות מודל תלת מימד של חלק מהמשחק אותו רוצים לבחון. למעט ספורט, ניתן לראות שימושים לטכנולוגיות אלו בתחומים נוספים - ברפואה, בתחום המכוניות האוטונומיות, באדריכלות ועיצוב פנים וכו'.

ישנן מספר דרישות עיקריות כאשר באים לבחור מימוש לפתרון לבעיית השחזור:

1. אמינות ודיוק: כמה תוצאות המודל דומות לעולם האמיתי. בפועל, זהו מדד סובייקטיבי ולכן נרצה להשתמש במדד כמותי. במושגים של מערכות לומדות, נרצה לבחור פונקציית מחיר מתאימה ליישום.
2. זמן: ישנם יישומים בהן נרצה לקבל תוצאות תוך זמן קצר, כמו בדוגמה של משחק הכדורגל וישנם יישומים שאין הגבלה של זמן, לדוגמה יצירת מודל תלת מימדי של הקוליסאום. נשים לב שיש מודלים שניתן לאמן מראש (בד"כ זמן ארוך) אך לאחר מכן השימוש בהם יעיל בזמן.
3. עמידות: בעולם האמיתי, התמונות והמידע שאנחנו מכניסים למערכת הוא "רועש". לכן נרצה מודל שיהיה עמיד לרעש ולסוגים שונים של שגיאות ועיוותים שיכולים להתרחש.

4. כוח חישובי: רוב היישומים מוגבלים בכוח חישובי ולכן נצטרך להתחשב בכך, כמו שגם ריצה ארוכה של תוכנה צורכת חשמל רב ובכך פוגעת בסביבה.

5. כסף:

ישנם שיקולים רבים נוספים וישנו Trade-off בין שיקולים אלו. בפרויקט זה ניסינו לשפר את עמידות המודל על ידי ביצוע Data Augmentation בשלב האימון.

### 3. NeRF - Representing Scenes as Neural Radiance Fields for View Synthesis<sup>1</sup>

נתאר בקצרה את האלגוריתם:

שלב האימון:

כניסה : המיקום היחסי של המצלמה  $(x,y,z)$  והזווית המרחבת של המצלמה  $(\phi, \theta)$ .

שלב ה- **Forward** : הקואורדינטות ה-5 מימדות נראות לרשת נוירונים רב שכבתית (MLP) שמוצאה היא צפיפות וצבע במרחב של הסצנה. באמצעות כלים של volume rendering מופקת תמונה באמצעות המידע שחושב בשלב קודם. פונקציית המחיר: פונקציית המחיר מחושבת ע"י הסכום הכולל של השגיאות בריבוע:

$$\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}} \left\| \hat{C}(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2$$

where  $R$  is the set of rays in each batch, and  $C(\mathbf{r})$ ,  $\hat{C}(\mathbf{r})$  are the ground truth and predicted RGB colors for ray  $\mathbf{r}$ .

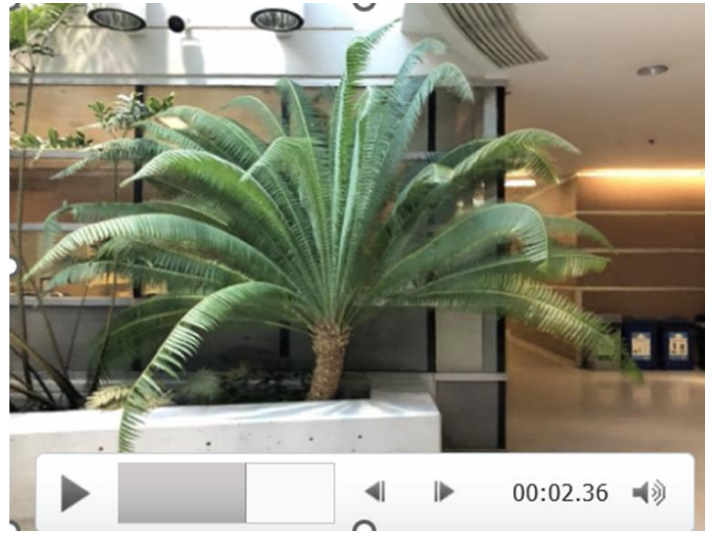
לאחר מכן משתמשים בכלים של אופטימיזציה על מנת לאמן את ה- MLP. השתמשנו ב Adam Optimization, שיטה נפוצה מאוד בתחום של למידה עמוקה.

שלב החיזוי: בהינתן קואורדינטות מרחביות, ניתן ליצור תמונה חדשה של האובייקט עם עומק שנלקחו מאותו מיקום מצלמה וזוויות מרחביות שניתנו למערכת כקלט.

---

<sup>1</sup> Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." *Communications of the ACM* 65.1 (2021): 99-106.

ניתן לראות באיור 2, את תוצאות המודל כפי שאומנו במחשבים של המעבדה:

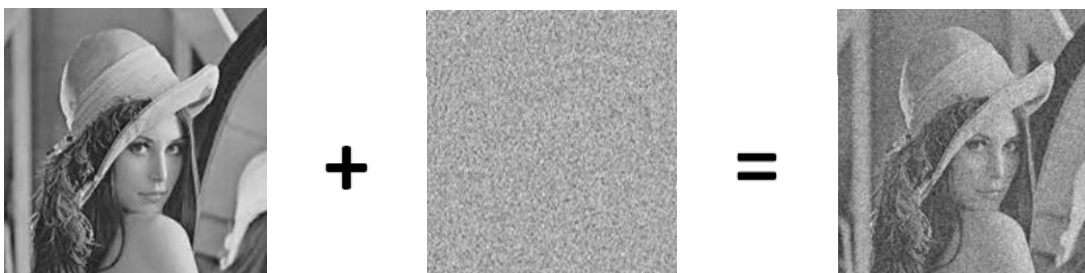


אפשר לראות שקיבלנו שחזור תלת ממדי איכותי מאוד כאשר רואים את כל הפרטים בבירור, גם הפרטים עם טקסטורה עדינה מאוד כמו העלים של העץ. כמו כן, רואים צבעים הדומים מאוד לצבעים האמיתיים ובגוונים שונים של אותו הצבע בדיוק כמו בחיים האמיתיים.

## 4. ניסיונות לשיפור המודל:

כדי לשפר את התוצאות, רצינו שהמודל יהיה עמיד לוריאציות שונות של תמונות הקלט. לצורך כך, ביצענו שני סוגים של Data Augmentation בשלב האימון באמצעות שימוש ב TensorFlow ובדקנו איך זה משפיע על התוצאות.

הוספת רעש גאוסי עם ערכים שונים של אנרגיה (שונות)



להוספת רעש במהלך האימון של מודל מבוסס רשת לומדת יש אפקט רגוליזציה המשפר את מידת העמידות (robustness) של המודל וזה מעולה כי כפי שציינו מקודם בפרק המוטיבציה, בעולם האמיתי, התמונות והמידע שאנחנו מכניסים למערכת הוא לרוב "רועש". לכן נרצה מודל שיהיה עמיד לרעש ולסוגים שונים של שגיאות ועיוותים שיכולים להתרחש.

כמו כן, מחקרים רבים ציינו כי הוספת כמויות קטנות של רעש קלט לנתוני האימון מסייעת לרוב להכללה ולסבילות לשגיאות.



## היפוך תמונה



צורה נוספת בה השתמשנו לביצוע Data Augmentation בשלב האימון היא היפוך תמונה. ביצענו אוריינטציה של שיקוף אשר דומה להיפוך שקורה כשאנחנו מצלמים "סלפי".

בכך, אנחנו נותנים למודל דוגמאות "שקריות" שכן לא סביר שטרנספורמציה זו תקרה בעולם האמיתי מלבד תמונת "סלפי". למרות זאת, דוגמאות אלו הן בעלות אותן תכונות כמו התמונות האמיתיות.

## 5. מסקנות ורעיונות להמשך

לפי הממצאים שקיבלנו, ניתן להסיק כי המודל המוצג במאמר עמיד לרעש (גאוס) אך לא עמיד לשינויים קיצוניים בזווית של המצלמה כמו למשל היפוך של 180 מעלות כפי שביצענו בניסיונות שלנו לבדוק כיוונים לשיפור השיטה. מעניין לבדוק השפעת סוגים אחרים של Data Augmentation בשלב האימון על התוצאות כמו חיתוך תמונה, שינוי גודל תמונה, סיבוב בזוויות קטנות ועוד.

כמו כן, לדעתנו, שווה לחקור לעומק יותר איך ניתן לקצר את זמן האימון. כרגע אורך זמן הריצה הוא בערך 15 שעות מה שהקשה עלינו מאוד להגיע לשיפורים משמעותיים ולחקור כיוונים אחרים בזמן הניתן לנו.

בנוסף, מעניין יהיה להתמקד בתחום אחד ספציפי ולדאוג שהתוצאות של השחזור התלת ממדי בשיטה המוצעת במאמר יענו לצרכים של אותו התחום. למשל, בתחום של רפואה, במקרים רבים מייצרים צורות עצם תלת מימדיות להדמיה ולתכנון לפני ניתוח. נתוני צורות אלה מופקים ממערכי נתונים נפחיים המתקבלים ע"י חיישנים תלת ממדיים, כגון טומוגרפיה ממוחשבת (CT) והדמיית תהודה מגנטית (MRI). שיטה קונבנציונלית זו היא אינטנסיבית מאוד גם מבחינת זמן וגם מבחינה כספית. מעניין יהיה לבדוק את תוצאות השחזור של השיטה המוצעת במאמר כאשר נספק לה מספר תמונות של רנטגן. במידה והתוצאות יהיו ברמה סבירה, אפשר יהיה לבדוק כיוונים אפשריים לשיפור ובכך לקבל שיטה יעילה וחסכונית יותר ליצירת צורות עצם תלת מימדיות ובהמשך אפשר יהיה להרחיב את השיטה ולייעל אותה עבור יישומים רפואיים שונים אחרים. להלן מאמר מעניין אחר בנושא הזה של חוקרים ממכון ויצמן:

*Kasten, Y., Doktofsky, D., & Kovler, I. (2020, October). End-to-end convolutional neural network for 3D reconstruction of knee bones from bi-planar X-ray images. In International Workshop on Machine Learning for Medical Image Reconstruction (pp. 123-133). Springer, Cham.*

## 6. סיכום הפרויקט ורפלקציה אישית:

1. עבדנו עם מאמרים עדכניים מהשנים האחרונות בתחום שעדיין גדל ומשתפר כיום. כמו כן, התחום של שחזור תלת-מימד על ידי למידה עמוקה אינו נפוץ בטכניון.

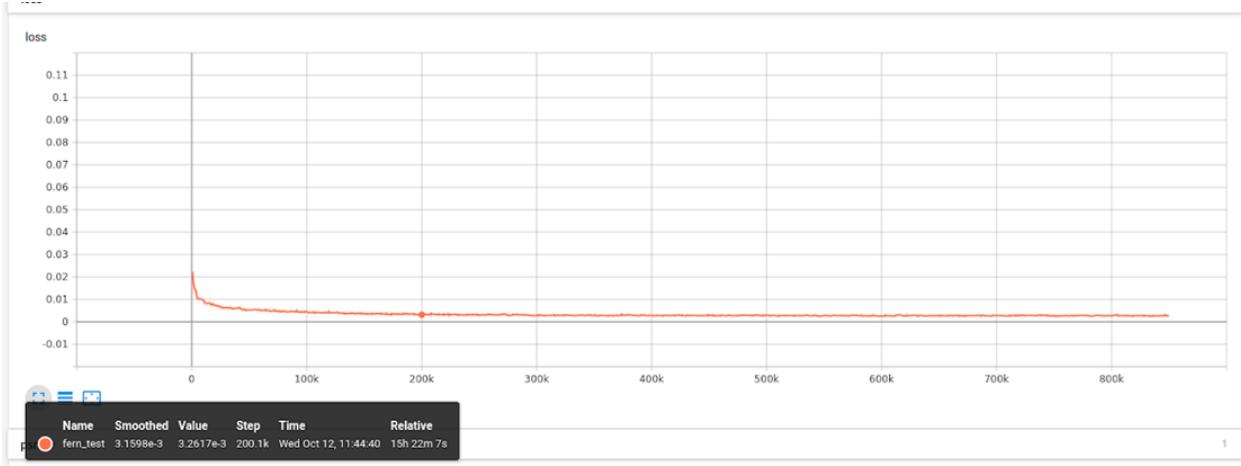
2. הפרויקט דרש עבודה הנדסית מקיפה: סקירת ספרות, קריאת מאמרים וניסיונות להריץ ולשפר קוד קיים באינטרנט, הבנת הצורך והאתגרים בתחום, בחינת אפשרויות לטיוב המודל וכו'.

3. כמו כל עבודה בתחום של מערכות לומדות, משאבים חישוביים וזמן הם צורך קריטי להצלחת הפרויקט. בהינתן מאגר מצומצם של שניהם, לדעתנו הצלחנו לקדם את הפרויקט בצורה יפה.

4. תהליך הלמידה שנעשה בפרויקט הינו עצום - מלמידה של הפלטפורמות השונות (tensorflow, torch) ללמידה של הרקע האקדמי למאמרים בתחום זה.

## נספחים:

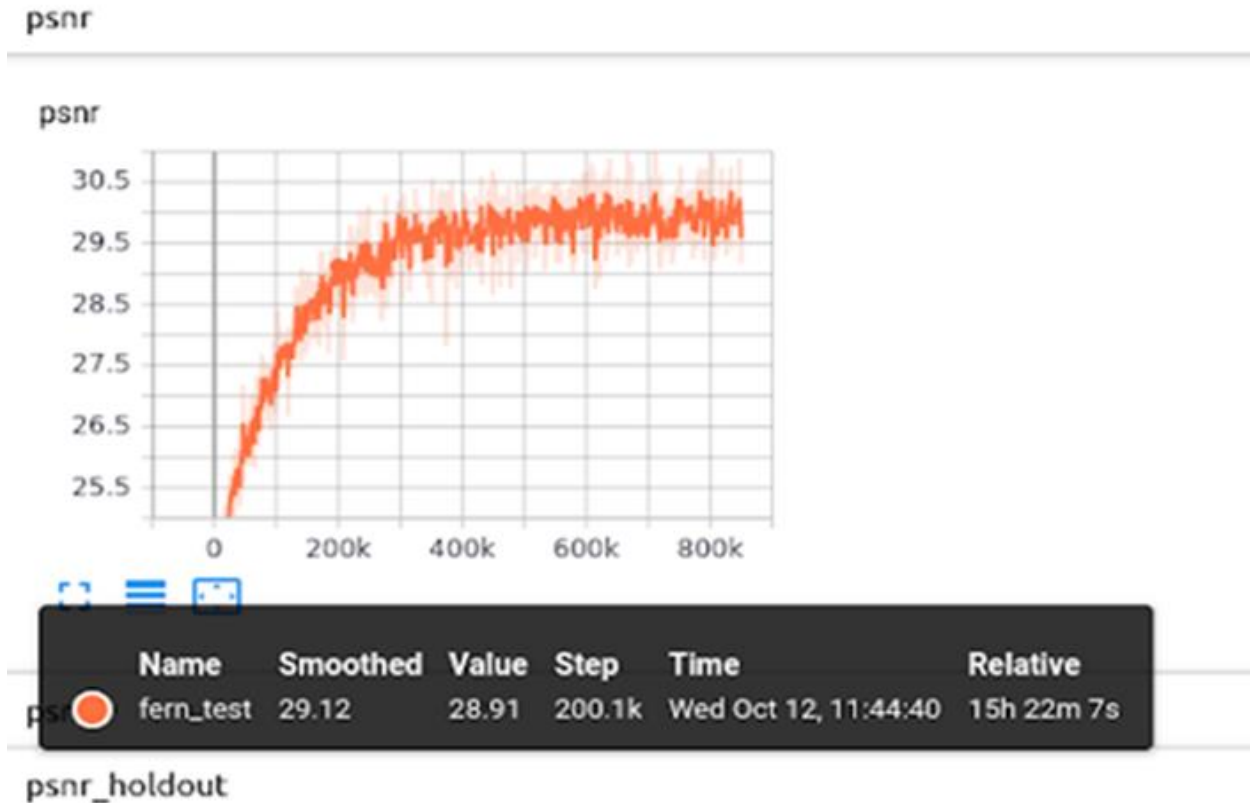
גרף המייצג את פונקציית המחיר בלי הוספת רעש:



Name	Smoothed Value	Value	Step	Time	Relative
fern_test	3.1598e-3	3.2617e-3	200.1k	Wed Oct 12, 11:44:40	15h 22m 7s

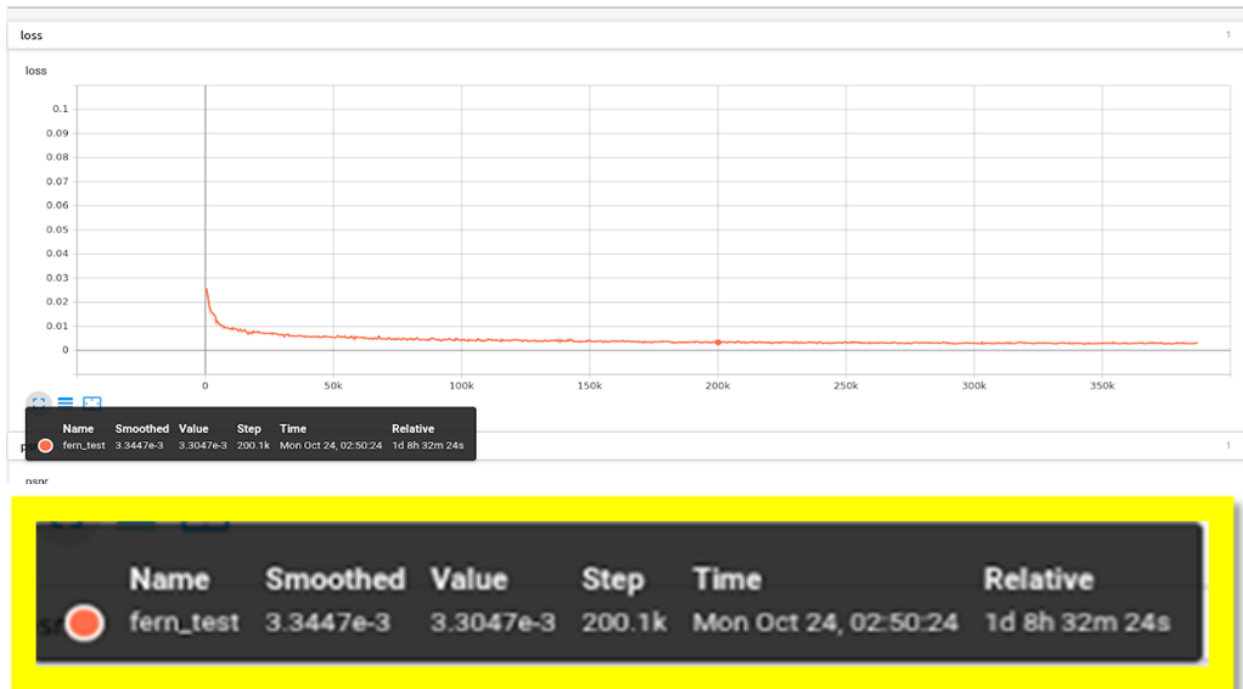
ניתן לראות בגרף את הקשר בין מספר האיטרציות של הריצה לבין ערך השגיאה לפי פונקציית המחיר. רואים שמתחילים מערך שגיאה גדול וככל שאנחנו מתקדמים עם מספר האיטרציות, ערך זה קטן. אפשר לשים לב כי לאחר K200 איטרציות, אנחנו "מתכנסים" ומתייצבים על אותו ערך של שגיאה אשר מופיע בצילום המצורף.

גרף המייצג את מדד ה **PSNR** בלי הוספת רעש:



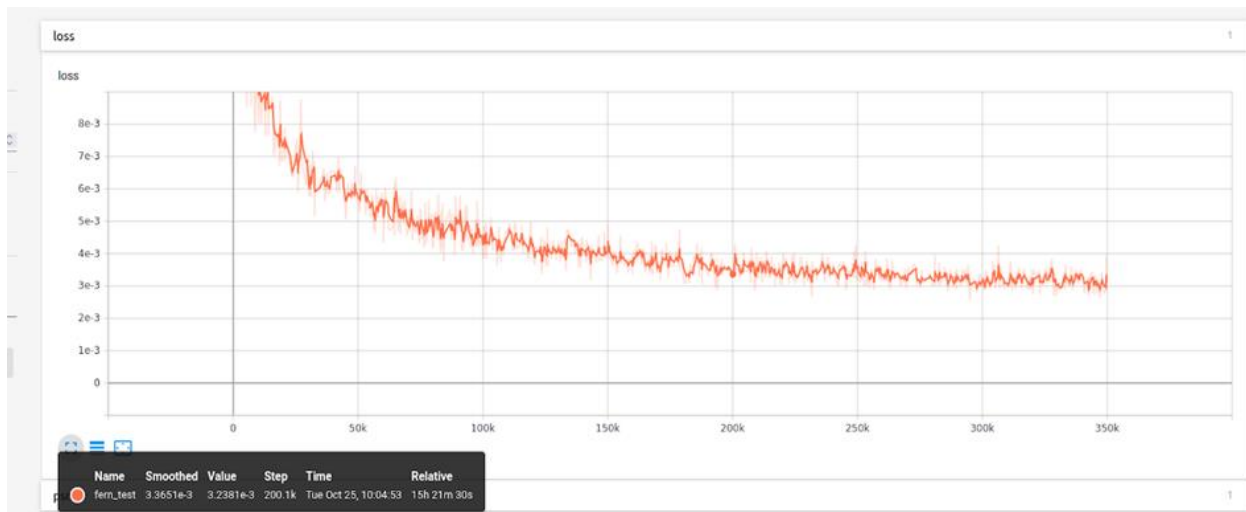
אפשר לראות בגרף את הקשר בין מספר האיטרציות לערך ה PSNR אשר מתאר את היחס בין עוצמת האות האמיתית לבין עוצמת הרעש. ערך יחס שהוא גדול מעיד על עוצמת אות גדולה יותר מעוצמת הרעש. ניתן לשים לב שלאחר K200 איטרציות, אנחנו מגיעים בערך לערך המקסימלי ומתחילים להתייבב. זה קורה לאחר 15~ שעות. אפשר לראות את ערך היחס לאחר K200 איטרציות בצילום המצורף.

גרף המייצג את פונקציית המחיר עם הוספת רעש גאوسی (שונות 0.001)



ניתן לראות בגרף את הקשר בין מספר האיטרציות של הריצה לבין ערך השגיאה לפי פונקציית המחיר. רואים שמתחילים מערך שגיאה גדול וככל שאנחנו מתקדמים עם מספר האיטרציות, ערך זה קטן. אפשר לשים לב כי לאחר K200 איטרציות, אנחנו "מתכנסים" ומתייצבים על אותו ערך של שגיאה אשר מופיע בצילום המצורף. כמו כן, אפשר להבחין בכך שערך השגיאה לאחר K200 איטרציות הוא מאותו סדר גודל של ערך השגיאה לפני הוספת הרעש.

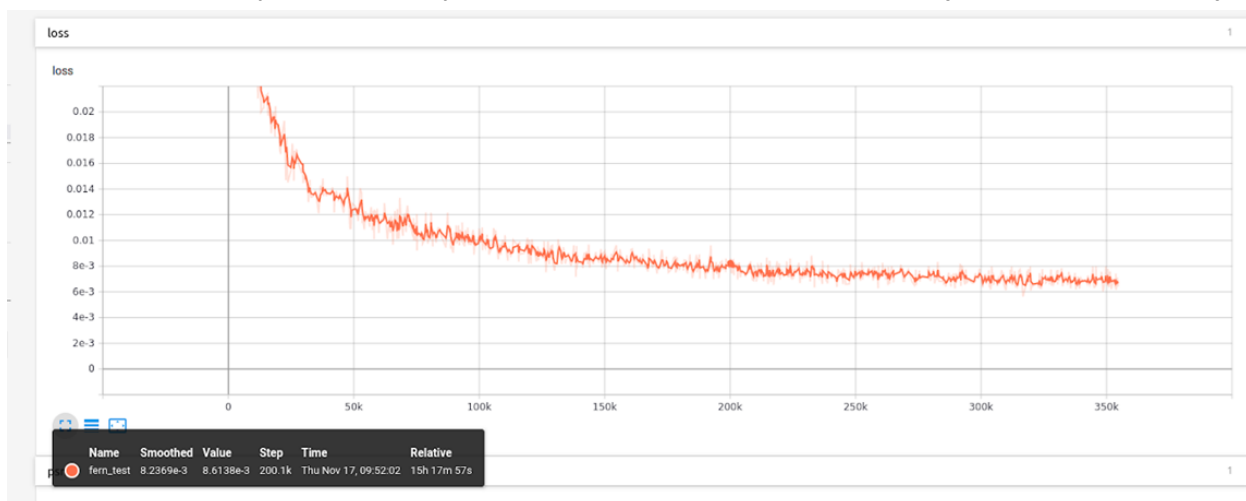
גרף המייצג את פונקציית המחיר עם הוספת רעש גאوسی (שונות 0.01)



Name	Smoothed Value	Value	Step	Time	Relative
fern_test	3.3651e-3	3.2381e-3	200.1k	Tue Oct 25, 10:04:53	15h 21m 30s

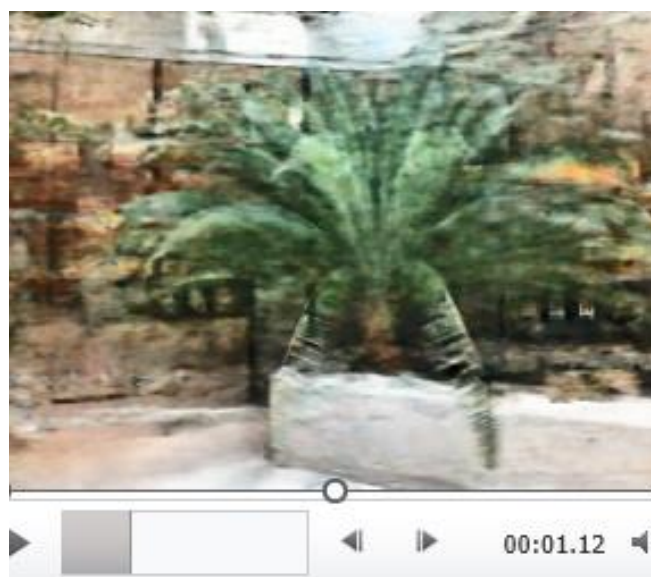
ניתן לראות בגרף את הקשר בין מספר האיטרציות של הריצה לבין ערך השגיאה לפי פונקציית המחיר. רואים שמתחילים מערך שגיאה גדול וככל שאנחנו מתקדמים עם מספר האיטרציות, ערך זה קטן. אפשר לשים לב כי לאחר K200 איטרציות, אנחנו "מתכנסים" ומתייצבים על אותו ערך של שגיאה אשר מופיע בצילום המצורף. כמו כן, אפשר להבחין בכך שערך השגיאה לאחר K200 איטרציות הוא מאותו סדר גודל של ערך השגיאה לפני הוספת הרעש.

גרף המייצג את פונקציית המחיר לאחר סיבוב תמונה (180 מעלות)



	Name	Smoothed	Value	Step	Time	Relative
ts	fern_test	8.2369e-3	8.6138e-3	200.1k	Thu Nov 17, 09:52:02	15h 17m 57s

ניתן לראות בגרף את הקשר בין מספר האיטרציות של הריצה לבין ערך השגיאה לפי פונקציית המחיר. רואים שמתחילים מערך שגיאה גדול וככל שאנחנו מתקדמים עם מספר האיטרציות, ערך זה קטן. אפשר לשים לב כי לאחר K200 איטרציות, אנחנו "מתכנסים" ומתייצבים על אותו ערך של שגיאה אשר מופיע בצילום המצורף. כמו כן, אפשר להבחין בכך שערך השגיאה לאחר K200 איטרציות הוא מאותו גדול יותר מערך השגיאה לפני הוספת הרעש יחסית בצורה משמעותית וזה בא לידי ביטוי בתוצאה של השחזור:



רואים שקיבלנו שחזור תלת ממדי לא ברור. מצד שני, אפשר עדיין להבחין בעץ ולהבין מה היינו אמורים לקבל.