

פרויקט א'

Further Steps in Precise Shape Completion

צעדים נוספים בשחזור מדויק של צורות

מגישים:

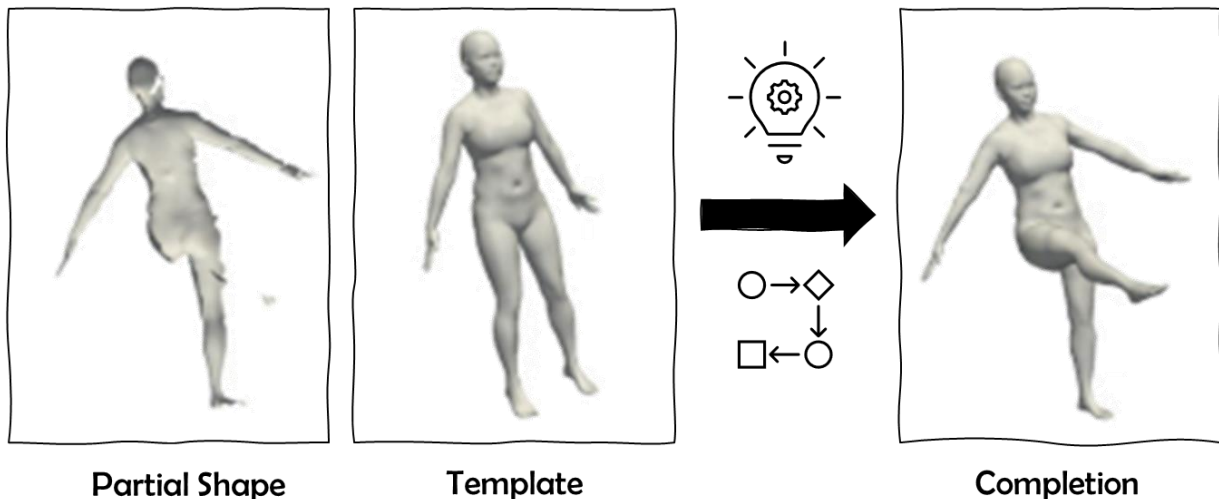
הדס רומוב

יפתח אדלשטיין

מנחה:

עידו עמנואל

סמסטר: אביב 2022



תוכן עניינים

3	מטרת הפרויקט	1
4	מבוא	2
4	תקציר	2.1
4	צבירת ידע	2.2
5	סקירת ספרות	2.3
	מאמר 1 - PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation	2.3.1
5		
6	מאמר 2 - Towards Precise Completion of Deformable Shapes	2.3.2
6	מאמר 3 - Implicit Neural Representation with Periodic Activation Functions	2.3.3
7	תיאור מפורט	3
7	סביבת העבודה	3.1
7	מערכי נתונים	3.2
7	הגדרות בסיסיות	3.3
7	בדיקת היתכנות	3.4
8	מערכת ניטור ומדידה	3.5
8	מבנה הרשת ההתחלתית	3.6
9	תוצאות ראשוניות	3.7
12	מימוש הרשת החדשה	4
12	רעיונות מרכזיים	4.1
12	ביסוס מתמטי	4.2
13	תוצאות	5
13	שיפור פונקציית ההפסד – התחשבות בהפסדי נפח :	5.1
20	שינויים ארכיטקטוניים – החלפת פונקציות האקטיבציה בפונק' סינוס :	5.2
23	ניסויים נוספים – ארכיטקטורת LSTM	5.3
27	אתגרים ופתרונות	6
28	סיכום ומסקנות	7
28	תודות	8
29	נספחים	9
29	9.1 מרכזו של הדמויות כלפי מרכז המסה	9.1
30	שרטוטים ואיורים	10
30	גרפים	11
30	טבלאות	12
31	ביבליוגרפיה	13

1 מטרת הפרויקט

לאור ההתפתחויות הטכנולוגיות בעולם המודרני, ובעקבות מציאות נגיף הקורונה שהתבססה בחיינו, בקרב אנשים רבים החל החיפוש אחר דרכים חדשות להתחבר זה עם זה ולשמור על קשר. היינו רוצים להיעזר במציאות וירטואלית בתור הבסיס לחיבור זה, מתוך שאיפה להיות מסוגלים להופיע בעולם וירטואלי יחד עם האנשים הקרובים לנו. שאיפה זו מלווה בקשיים – היכולת לשרוק את גוף האדם ולתפוס את תנועותיו והבעותיו זו משימה מסובכת, המוגבלת על ידי ציוד הצילום, הזווית והמיקום לצד קשיים נוספים. סריקה תלת-מימדית של אדם תדרוש צילום ממספר רב של מצלמות, על מנת להתגבר על מגוון בעיות הסתרה – ותדרוש ציוד יקר ומורכב לתחזוקה. בנוסף, אנו היינו מעוניינים בסריקה מדויקת, שנראית כמו בן האדם, ולא סריקה חלקה שרק תתאר את תנועתו והמחוות שהוא מבצע. הפתרון שאנו מציעים יאפשר צילום ממצלמה אחת, מזוויות שונות, ולשם כך נרצה להשלים סרטונים של סריקות חלקיות לכדי סרטון של הסריקה המלאה.

2 מבוא

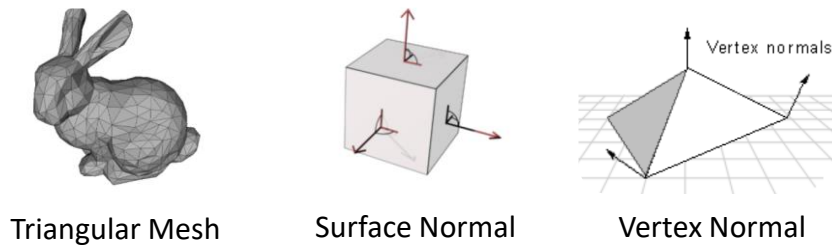
2.1 תקציר

במסגרת הפרויקט ננסה להרחיב את העבודה המתוארת במאמר "Towards Precise Completion of Deformable Shapes"⁵. ננסה לשפר את המודל הקיים של הרשת, תוך שינוי פונקציית ההפסד, ארכיטקטורת הרשת ותהליך עיבוד הנתונים.

2.2 צבירת ידע

כחלק מתהליך ההכנה לקראת העבודה בפרויקט, נדרשנו לרכוש מושגים והגדרות בתחום העיבוד הגאומטרי וניתוח של צורות. לשם כך למדנו באופן עצמאי קורסים רלוונטיים בתחום, המועברים על ידי קינאן קריין מאוניברסיטת קרנגי מלון¹, וג'סטין סולומון מ-MIT². במסגרת הלמידה הכרנו את המושגים הבאים:

1. Triangular Mesh - סוג של יריעה פוליגונית בגרפיקה ממוחשבת. היא מורכבת מאוסף משולשים (לרוב בתלת מימד) המחוברים בצמתים או צלעות משותפות.
2. Surface Normal - וקטור המאונך למישור המשיק למשטח.
3. Vertex Normal - הממוצע המשוקלל של הנורמלים המשטחיים המתאימים לפאות המכילות את הקודקוד הרלוונטי



איור 1

בנוסף למדנו אודות מונחים הקשורים לגאומטריה של משטחים המיוצגים באופן בדיד ורציף, ומושגים מעולם הטופולוגיה במרחב. כמו כן, מימשנו כלי לויזואליזציה של המשטחים והנורמלים לקודקודים, באמצעות ממשק MeshLab³¹:

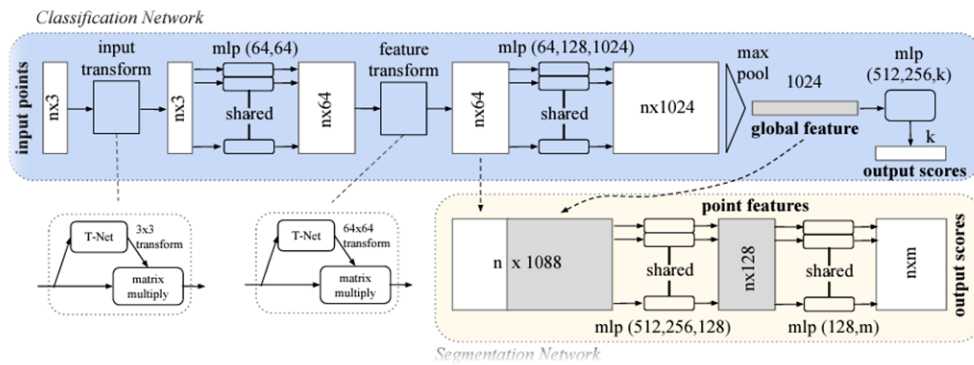


איור 2

¹ MeshLab הוא מערכת open source המיועדת לעיבוד ועריכה של יריעות mesh בתלת מימד.

2.3.1 מאמר 1 - PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation⁴

- המאמר עוסק בלמידה עמוקה מתוך ענני נקודות. מדובר בתהליך מסובך ולא מאוד יעיל באמצעות הדרכים הסטנדרטיות הנהוגות עבור תמונות, טקסט או אודיו, מאחר שענני נקודות אינם רגולריים בצורתם- כלומר אין חשיבות לסדר בתוך המבנה נתונים.
- ארכיטקטורות קיימות נוספות הכוללות קונבולוציות תלת מימדיות בסיסיות או RNNs המייצגים את הנקודה כמפה sequential, לא מספקות את הביצועים הדרושים. לכן, המאמר מציע ארכיטקטורה חדשה. הארכיטקטורה כוללת את הרכיבים העיקריים הבאים:
- א. Encoder (מקודד): המחלף מאפיינים בעלי משמעות מענן הנקודות באמצעות ארכיטקטורת MLP. החילוף כולל הן מאפיינים נקודתיים, והן מאפיינים גלובליים, המשורשרים לכל אחת מהנקודות.
 - ב. מטריצת טרנספורמציה המנסה לפצות על סטיות אפשריות בנתונים.
 - ג. Decoder (מפענח): שמשמש בצירוף של מאפיינים מקומיים וגלובליים לטובת סגמנטציה של חלקים שונים בענן או סיווג של הענן כולו.

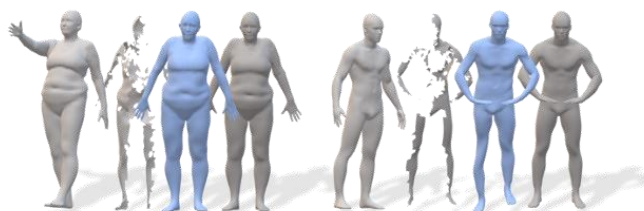


איור 3

2.3.2 מאמר 2 – Towards Precise Completion of Deformable Shapes⁵

המאמר עוסק בהשלמה מדויקת של צורות – וספציפית, בני אדם. הכוונה בהשלמה מדויקת, היא בניגוד להשלמה מבוססת מודל פרמטרי, שתיצור צורה חלקה, ופשוטה, שתתאר את התנועה בצורה מדויקת, אך לא את הפרטים העדינים של הבן אדם.

המאמר הזה שואף לבצע את ההשלמה בצורה משמרת פרטים – תווי פנים, אצבעות ידיים. המאמר מתאר ארכיטקטורה חדשה, שמשמשת ב-PointNet בתור מקודד. לתוך מקודד זה, אנו מכניסים 2 ענני נקודות שונים. אחד מענני הנקודות הוא התמונה החלקית ממנה התחלנו, והשני ענן נקודות שלם של אותו בן אדם, אבל בפוזוה שונה. אלו דרישות סבירות ממערכת אמיתית, מאחר ובמצב זה, נבצע סריקה טובה ומלאה, חד פעמית של הדמות, ונשתמש בסריקה זו על מנת לשמר את הפרטים העדינים במהלך ריצה, בשיתוף עם סריקה חלקית ומהירה, מזווית צילום אחת. המקודד מייצר וקטורי מאפיינים, שנכנסים למפענת, שבמוצאו, מייצר השלמה יפה ואיכותית. מאמר זה הוא הבסיס של העבודה שלנו בפרויקט.

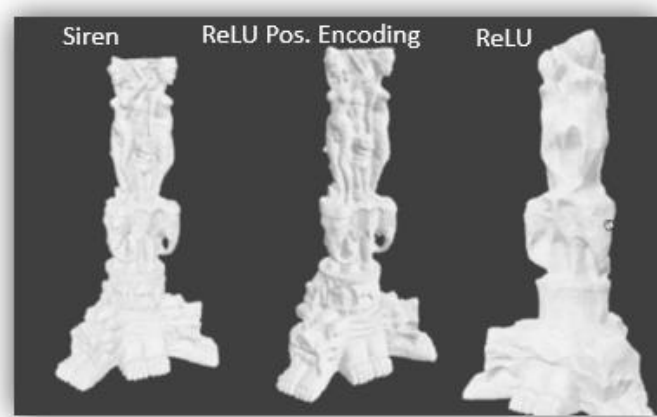


איור 4

2.3.3 מאמר 3 - Implicit Neural Representation with Periodic Activation Functions⁶

בשנים האחרונות, משתמשים ברשתות נוירונים על מנת לייצג מידע (ענני נקודות, תמונות, סרטונים). רשתות אלו, מייצגות מידע בתדרים נמוכים וגבוהים כאחד, וחוות קשיים בייצוג תדרים גבוהים. על מנת להתמודד עם בעיה זאת, רשתות הוסיפו שכבות של Positional Encoding² שעזרו בבעיה, אך מאמר זה הציע גישה יותר אפקטיבית(ניסיונית)- להחליף את כל הפונקציות אקטיבציה ברשת לסינוסים, ושינוי אתחול המשקלים הראשוני של הרשת בהתאם.

במאמר נתנו דוגמאות בעולם של תמונות, סרטונים, שמע, ענני נקודות ופתרון מד"חיס, מאפשרת לבצע גזירה של הרשת בצורה מאוד פשוטה – נגזרת של סינוס, היא גם כן סינוס(עם הפרש פאזה), מה שבמצבים מסוימים מאפשר לנו לפתור מד"ח בצורה פשוטה, ומקל על חישובי back-propagation.



איור 5

² שכבה, שמכפילה את הכניסה בסינוסים בתדרים שונים, ומעלה בכוח את תדרי המערכת. נקרא גם fourier features הוכח ניסיונית שעוזר

3 תיאור מפורט

3.1 סביבת העבודה

הפרויקט כתוב ברובו בשפת Python. במסגרת העבודה השתמשנו בחבילת ניהול הספריות conda, והתקנו מגוון ספריות המיועדות ללמידה עמוקה, עיבוד תמונות וייצוג נתונים – הבולטות שבהן הן pytorch, numpy, lightning. כמו כן, המרנו מערכת אמות מידה ומטריקות (benchmarking) משפת Matlab ל-Python, על מנת לאפשר מעקב מיטבי אחר ביצועי הרשת. לטובת הצגת התוצאות וניתוח ויזואלי מיטבי, נעזרנו בחבילת Wandb – weights & biases, המכילה ממשק אינטרנטי נגיש ונוח להצגת גרפים, טבלאות, מטריקות וניתוח נתונים עבור למידה עמוקה.

3.2 מערכי נתונים

לצורך האימון ובחינת המודל שלנו השתמשנו בשני מערכי נתונים:

- א. FAUST: מאגר המכיל סריקות של 10 דמויות שונות במגוון פוזות. כל סריקה בנויה מיריעת משולשים³ ושומרת על התאמה⁴ - כך שכל צומת שמייצג נקודה מסוימת בגוף הדמות ימצא באותו מיקום סידורי בכל ענן נקודות.
- ב. DFAUST: בנוסף למבנה של FAUST, DFAUST מכיל מספר משתנה של פריימים עבור כל פוזה, כך שלמעשה מתקבל מידע 4-מימדי. מעלה את נפח המידע, ונותן לנו עוד מימד של templates ומבנים לעבוד עליהם.

3.3 הגדרות בסיסיות

1. Full View – ענן נקודות מתוך המאגר, כאשר אנו רואים את כל הסריקה של הבן אדם.
2. Partial View – חיתוך של ענן הנקודות, בהתאם לזווית צילום מדומה כלשהי, כאחד מ-10 נקודות מבט שונות (כאשר אנו מחלקים את ה-360 המעלות, ל-10 זוויות צילום דיסקרטיות)
3. Template – סריקה מלאה, מפוזה שונה, מתוך המאגר, שמתפקדת בתור Prior עבור הרשת, ונותנת לה מידע על איך נראה הבן אדם, תחת ראייה מלאה, לצורך שחזור
4. Reconstruction – בהינתן Partial View ו-Template, התוצר שנוצר מהרשת. משווה ל-Full View ממנו נוצר ה-Partial View
5. Ground Truth - Full View ממנו נוצר ה-Partial View

3.4 בדיקת היתכנות

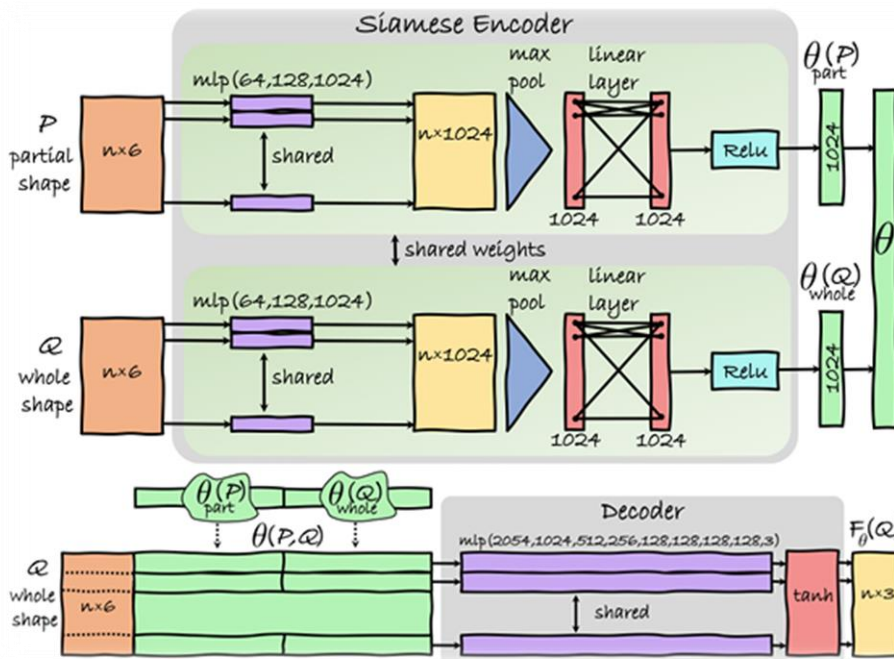
השתמשנו בארכיטקטורה קיימת לבדיקה של שחזור הסרטונים, וראינו שהתהליך אפשרי - אך מייצר תופעות בלתי רצויות, כגון: איבוד נפח וניפוח יתר של איברי גוף, מריחה, חוסר חלקות בתנועה, אפקטים של jitter ועיוותים בגוף הדמות. בנוסף, יש לציין כי הארכיטקטורה הקיימת נבנתה להתמודד עם סריקות יחידות בזמן, כמו תמונות שנלקחו במנותק מהסביבה - ולכן במהותה הארכיטקטורה לא בנויה להתמודד עם סרטונים כראוי. לדוגמא, היא דרשה לספק תבנית מסוימת על מנת לבנות השלמה כנגדה.

triangular mesh³
correspondence⁴

3.5 מערכת ניטור ומדידה

לצורך הצגה מדויקת ואמינה של התוצאות שקיבלנו תוך כדי העבודה, השתמשנו בفلטפורמה בשם (Weights & Biases) WandB, המאפשרת מעקב וויזואליזציה של העבודה המתבצעת ושימושית מאוד בתחומי למידת מכונה. המערכת משלבת ספריית קוד בשפת פייתון, שבאמצעותה ניתן להגדיר מהם הפרמטרים, הטבלאות, הגרפים והתוצרים אותם נרצה להציג בממשק האינטרנטי. במהלך הפרויקט נדרשנו ללמוד לשלב את הפונקציות של WandB בתוך הקוד שלנו, לקבלת גישה נוחה לכל הנתונים.

3.6 מבנה הרשת ההתחלתית



איור 6

נתאר את מבנה הרשת לפי חלקים:

א. Encoder:

רשת MLP מבוססת PointNet – מקבלת 2 ענני נקודות – Partial View + Template, ומחלצת מהם מאפיינים גלובליים. מאפיינים גלובליים אלה יתארו לנו בעצם מה מעניין בתוך ענן הנקודות, עבור הרשת, כדי לפענח לאחר מכן את הצורה המלאה. כפי שמתואר בציור, משרשרים את הווקטורים הגלובליים שחולצו, ומשרשרים אותם לצורה במפענח.

ב. Decoder:

רשת MLP נוספת, שלוקחת את ענן הנקודות של התבנית, בציורף המאפיינים הגלובליים אותם מצאנו, ומייצרת מהם שחזור טוב ככל שיכלה של הסריקה המקורית.

3.7 תוצאות ראשוניות

בתור התחלה, ניסינו לשחזר את תוצאות המאמר "Towards Precise Completion of Deformable Shapes" ולהשוות את הערכים שהגענו אליהם עם הערכים המצופים בריצת baseline.

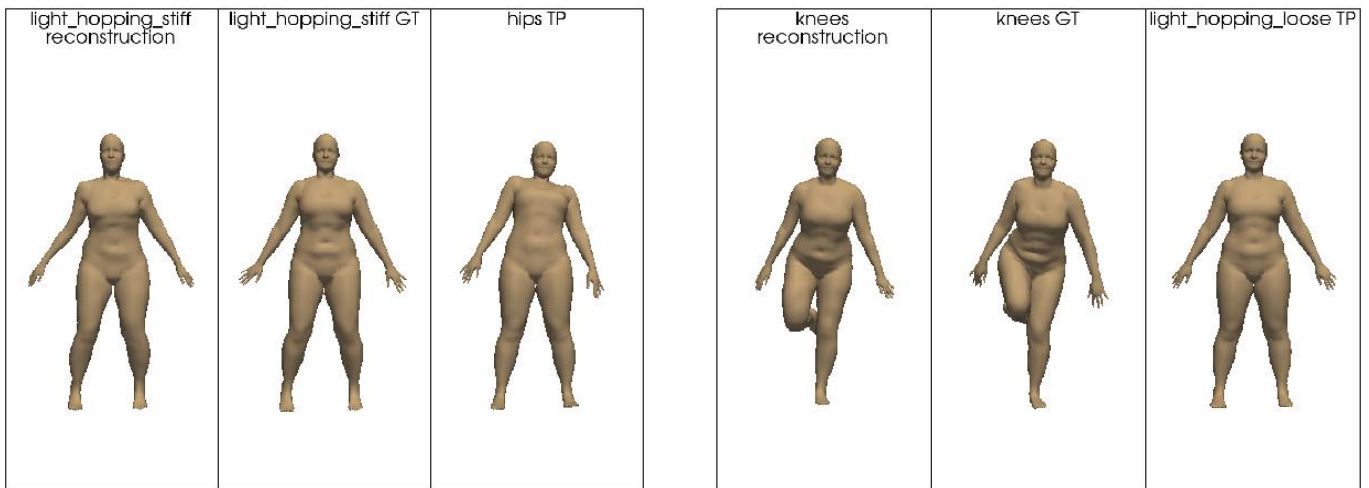
להלן חלק מהשחזורים הסטטיים שהצלחנו לקבל:

*עבור כל פוזה שניסינו לשחזר נציג את האלמנטים הבאים, מימין לשמאל:

א. TP – Template View

ב. GT – תוצאת ה-ground truth, כלומר התמונה האמיתית אותה ניסינו לשחזר.

ג. Reconstruction



איור 7

בנוסף לתוצאות הוויזואליות, מדדנו את ערכי ה-Loss המתקבלים באופן כמותי, עבור כלל המאגר, בשלושת שלבי הריצה של הרשת (Train, Validation, Test):

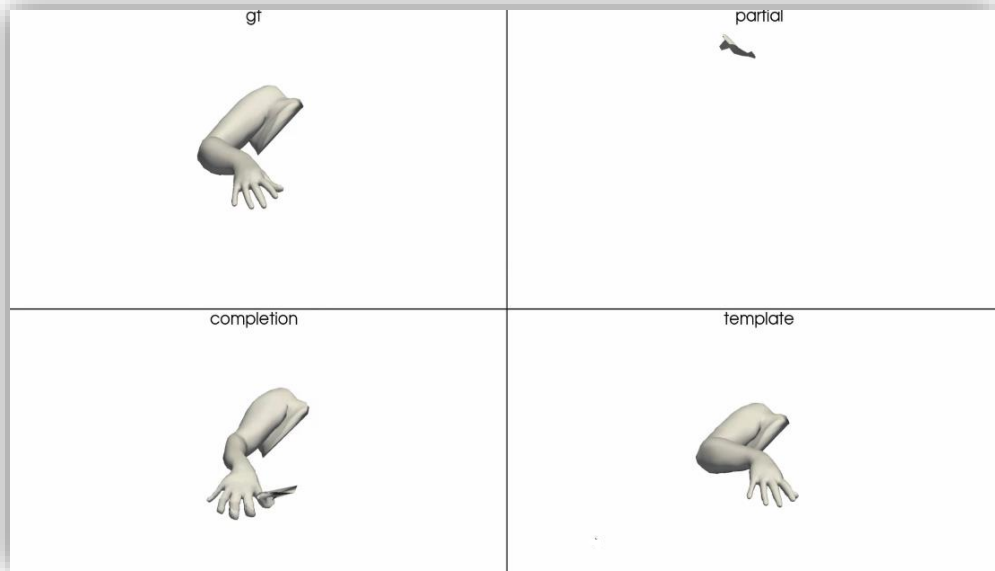
Stage	Completion to GT Vertex MSE	Mean L1 Completion to GT Volume Deformation [%]
Train	0.03735	5.278
Validation	0.03879	8.09
Test	0.04319	9.373

טבלה 1

מהתוצאות עולות מספר בעיות עיקריות הניתנות להבחנה :

א. הופעת שובל תנועה :

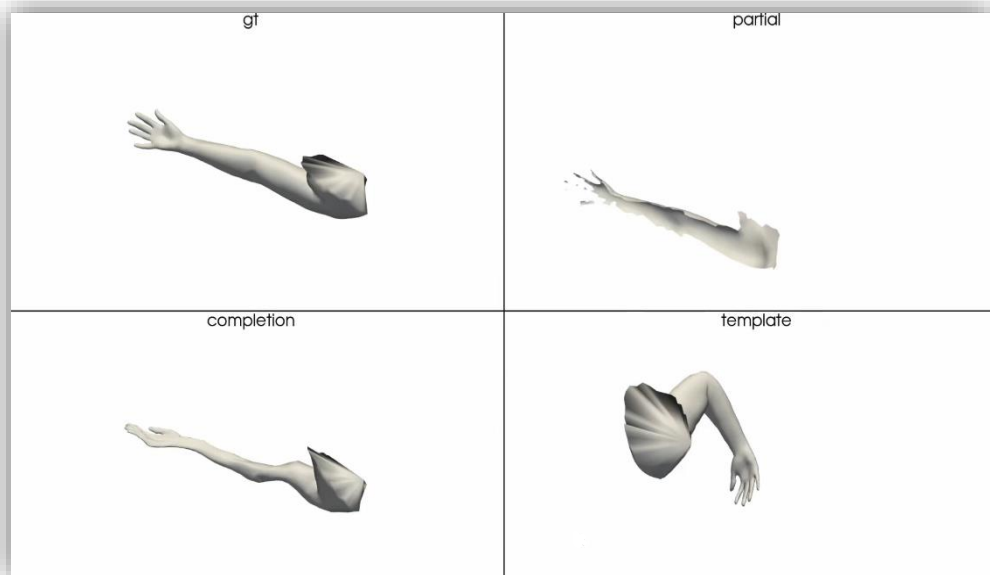
התופעה נצפתה בעיקר באזורי הגפיים, בהם התנועה לאורך הזמן הייתה יותר מורגשת. שובל התנועה נובע מכך, שבמסגרת השחזור, נקודות מסוימות חוו "הזחה", והשחזור, שניסה לשחזר אותן מנקודת המבט החלקית מיקם אותן הרחק משאר הנקודות.



איור 8

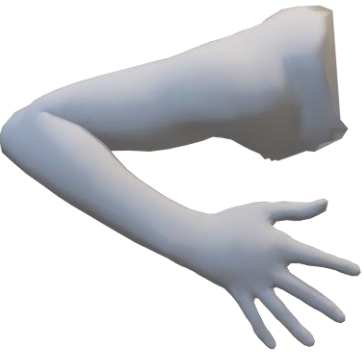



ב. אובדן נפח :

נראתה תופעה משמעותית של אובדן נפח, בעיקר באזורי הגפיים :



איור 9

ג. בעיות בשחזור רזולוציה:

<u>Ground Truth</u>	<u>Baseline Results</u>
	
	

איור 10

כפי שניתן לראות, הרשת לא הצליחה "לתפוס" כראוי אזורים המאופיינים בתדרים גבוהים, כלומר איברים בעלי צורה עדינה ומפורטת יותר – כמו האצבעות.

4 מימוש הרשת החדשה

4.1 רעיונות מרכזיים

על מנת להתמודד עם הבעיות שהוצגו לעיל, בחרנו להתקדם בשני כיוונים עיקריים:

א. שיפור פונקציית ההפסד:

שימוש מיטבי בתלות המרחבית והזמנית של ענני הנקודות.

האלמנט העיקרי שניסינו להוסיף לפונקציית ההפסד הוא שינויי נפח בענן הנקודות כולו, וכן שינויי נפח בכל איבר בגוף הדמות, כך שפונקציית ההפסד תתחשב בשינויים אלה ו"תעניש" איברים שהפער הנפחי בין תוצאת השחזור לתוצאת האמת בהם הוא גבוה במיוחד.

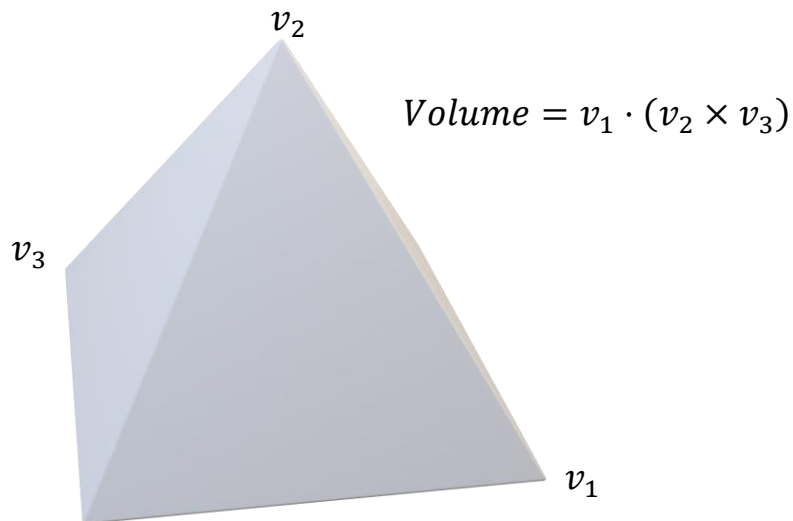
ב. שינויים ארכיטקטוניים:

החלפה של פונק' האקטיבציה של הרשת בפונקציות סינוס, בהתאם למאמר שהצגנו, וניסיון להחלפת הdecoder של הרשת LSTM Decoder, לשם שיפור החלקות הזמנית של הסרטונים המשוחזרים.

4.2 ביסוס מתמטי

נגדיר הפסד נפח באופן הבא:

ראשית, נחשב נפח פירמידה, כלפי ראשית הצירים. כלומר עבור כל יריעה משולשת, שמהווה חלק מהצורה, ניצור את הווקטורים התואמים לנקודה (0,0,0), ונחשב את הנפח לפי התרשים:



איור 11

כעת, עבור $f \in F$, כאשר F הם המשולשים המרכיבים את הצורה, נגדיר את $V_x(f)$ שיהיה הנפח בעל

הסימן, של ההשלמה, ו $V_y(f)$ נפח בעל הסימן של Ground Truth.

נגדיר את פונקציית ההפסד להיות:

$$L_{volumetric} = \sum_{f \in F} |V_x(f) - V_y(f)|_2^2$$

כאשר בשליטתנו קבוצת המשולשים עליה אנו מחשבים את הנפח. כלומר, אנו יכולים להטיל את ההפסד, רק על המשולשים המרכיבים את היד הימנית/שמאלית/רגליים/ראש או כל הגוף.

5.1 שיפור פונקציית ההפסד – התחשבות בהפסדי נפח :

הוספנו לפונקציית ההפסד התחשבות בנפח – כלומר, ניסינו "להעניש" את הרשת שלנו עבור שחזורים שבהם איבדה הדמות מנפחה.

תהליך זה התבצע במספר איטרציות, כאשר בחנו בכל פעם כיצד משפיעה ההתחשבות בנפח באזורים ספציפיים בגוף הדמות. למשל, עבור "מרכז הכובד" של הדמויות (אזור הטורסו) לא היה צורך בהענשה מיוחדת – מכיוון שאזורים אלה לא חוו תזוזה משמעותית במהלך התנועה, ולכן נפחם נשמר באופן יחסי. לעומת זאת, באזורי הגפיים, שהיו מאוד דינמיים במהלך האנימציה, נצפו עיוותים והפסדי נפח דומיננטיים – ולכן היה צורך לתת משקל רב יותר לשינויים באזורים אלה.

ביצענו 4 ניסויים הכוללים שימוש בהפסד נפח :

1. הפסד נפח על כל הגוף
2. הפסד נפח ביד ימין
3. הפסד נפח ביד שמאל
4. הפסד נפח בשתי ידיים

הפסד נפח בכל הגוף

בניסוי זה, ניסינו להעניש את ערך ה-Loss על שינוי בנפח הכולל של הדמות, כך שלמעשה הפונקציה מתעלמת מעיוותים אפשריים באיבר ספציפי. כתוצאה מכך התקבלו עיוותים עבור חלק מהדמויות, משום שפונקציית ההפסד החדשה לא הצליחה להתכנס כראוי.

הפסד נפח ביד ימין/שמאל

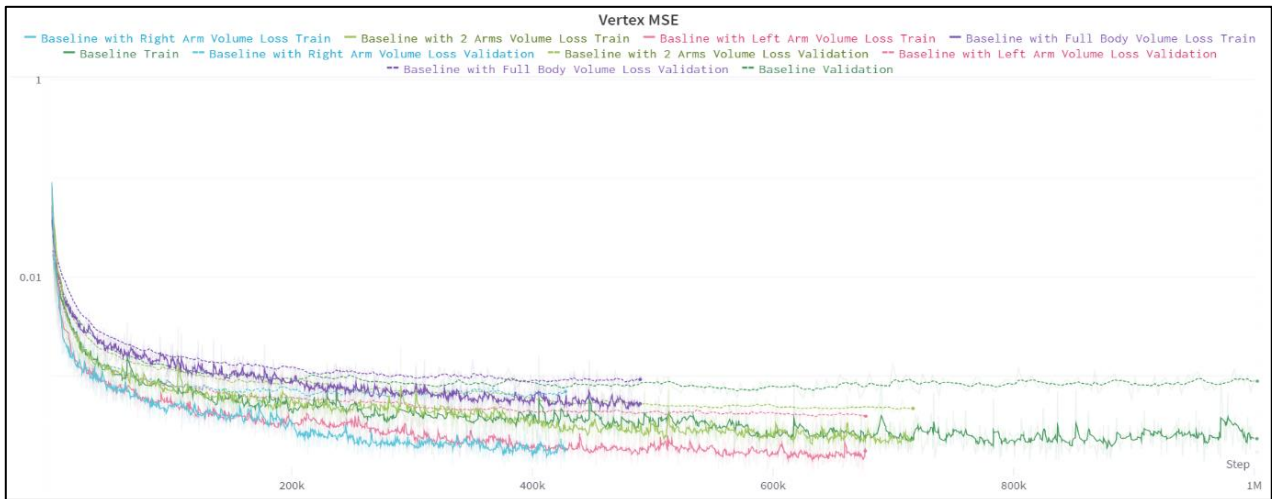
הניסוי כלל תוספת של הפסד נפח על אחת מהידיים (ביצענו אותו עבור כל יד בנפרד). ראינו שאכן האיבר שבחרנו להתייחס אליו שוחזר בצורה טובה יותר מבחינת שמירה על נפחו, ביחס לאיבר השני. כמו כן, קיבלנו שהוספת ה"הענשה" באיברים מסוימים פגעה בשגיאת הנפח הכולל והגדילה אותה, אך שיפרה את השגיאה עבור איבר ספציפי מתאים.

הפסד נפח בשתי הידיים

בניסוי זה הוספנו התחשבות בהפסד נפח בשתי הידיים. מטרתו היא בעיקר להראות שהשינוי משפיע על שתי הידיים במידה דומה.

השווינו בין שלבי האימון והוולידציה בכל אחד מהניסויים, וריכזנו את התוצאות בגרפים הבאים :

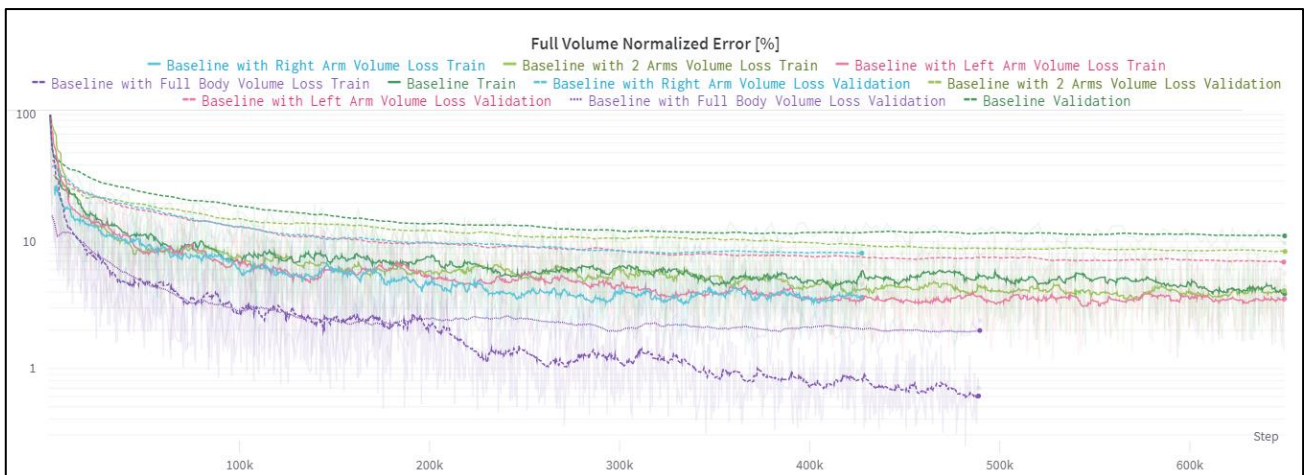
א. גרף של מדד MSE עבור הצמתים :



גרף 1

ניתן לראות שמדד ה-MSE המינימלי מתקבל עבור הניסויים בהם התווסף הפסד נפח על איבר בודד. לעומתם, דווקא בניסוי בו התחשבנו בהפסד הנפח הכולל, מדד ה-MSE יחסית גבוה. תוצאה זו ניתנת להסבר ע"י העובדה שהתקבל trade off בין המדדים השונים – כלומר, ככל שהשיפור הנפחי של כלל הגוף מהותי יותר, מדד ה-MSE מפצה על כך ומתקבלת תוצאה פחות טובה – אפילו ביחס לריצת ה-baseline.

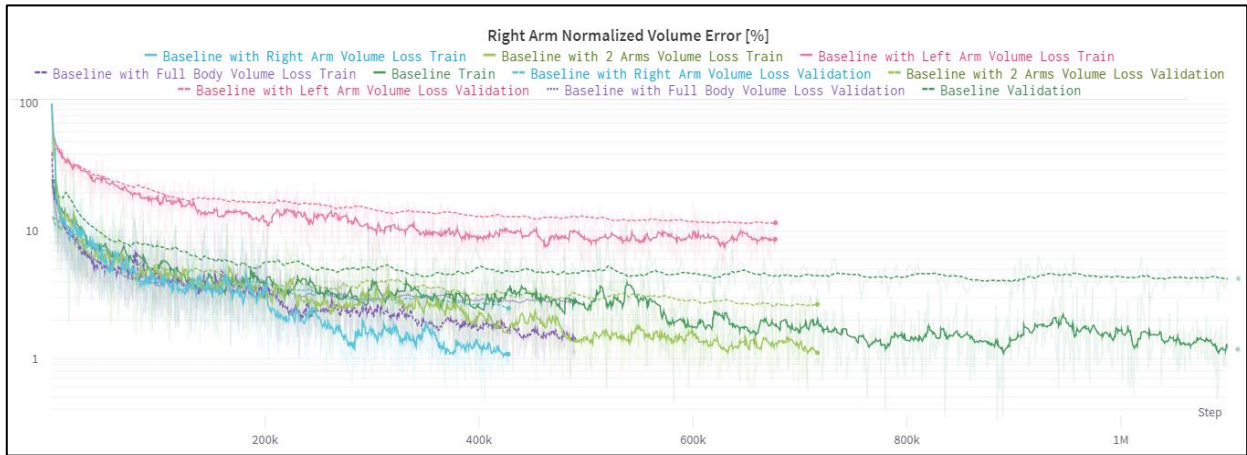
ב. גרף של שגיאה נפחית מנורמלת (עבור כל הגוף) :



גרף 2

בגרף זה, ניתן לראות בבירור ששגיאת הנפח הכולל המינימלית התקבלה בניסוי בו הוספנו הפסד נפח כולל- תוצאה שאכן ציפינו לקבל ומתיישבת עם השערותינו. כמו כן, השגיאה עבור כל אחת מהידיים דומה בקירוב, אם כי גבוהה יותר, משום שבניסויים אלה ה"עונש" הוגדר רק עבור איברים אלה. בהתאם למצופה, ריצת ה-baseline משקפת את הבעייתיות שבארכיטקטורה הקיימת מהבחינה הזו, ומגיעה לערך השגיאה הגבוה ביותר.

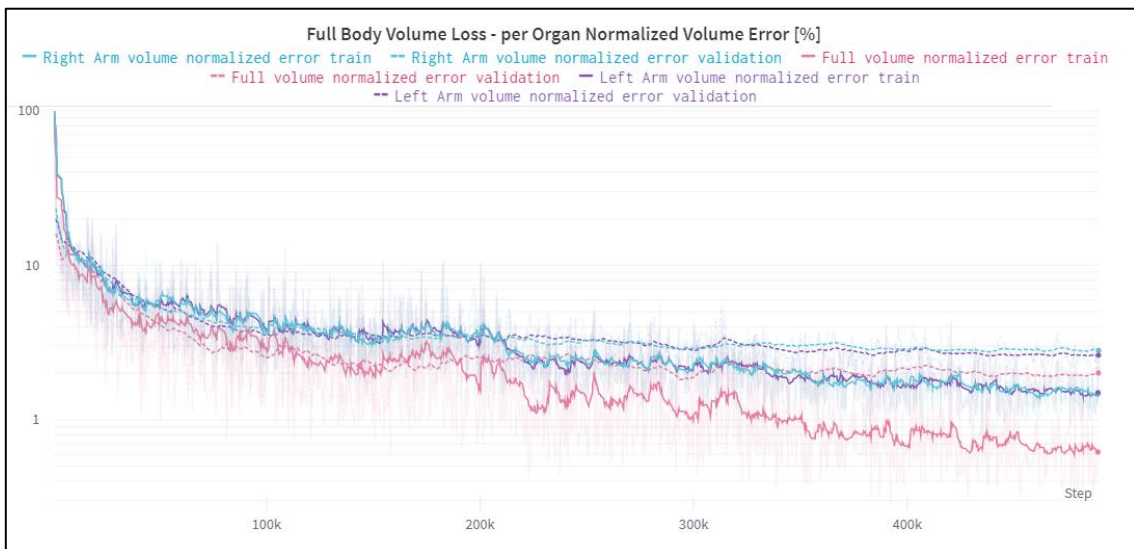
ג. גרף של שגיאה נפחית מנורמלת (עבור יד ימין) :



גרף 3

בחרנו להציג את השגיאה הנפחית המתקבלת עבור יד ימין, כדי להציג את ההבדל בין תוספת הפסד נפח ביד ימין לבין יד שמאל. ניתן לראות שבניסוי בו התחשבנו בעיקר בהפסד ביד שמאל, ומזערנו את השגיאה הכוללת על פיו, מדד הנפח ביד ימין לא הצליח להתכנס בצורה מספקת, גם לעומת ה-baseline. למעשה, ככל שנותנים יותר משקל ליד שמאל בפונקציית ההפסד, התוצאות עבור יד ימין פחות טובות. מכאן ניתן להסיק את המסקנה המתבקשת, שיש לוודא מהו האיבר הבעייתי מבחינת המדד הרלוונטי, ולהשתדל להתחשב בעיקר בו כאשר מבצעים שינויים בפונקציית ההפסד. כמו כן, השווינו בין השגיאות המתקבלות עבור ניסוי ספציפי – כלומר, עבור תוספת של הפסד נפח מסוים, בחנו את ההשפעה על מדדי השגיאה השונים שהתקבלו לאורך האימון :

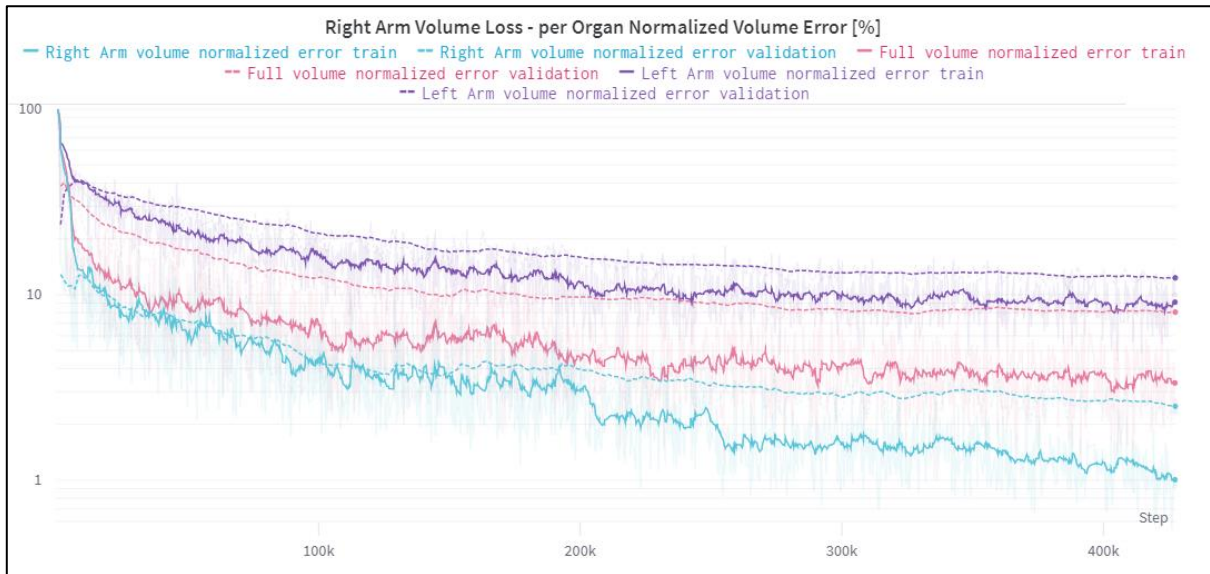
א. תוספת הפסד נפח בכל הגוף :



גרף 4

גם כאן ניתן לראות באופן ברור את הקשר בין השינוי בפונקציית ההפסד למזעור השגיאות השונות – בניסוי של הפסד נפח כולל, השגיאה הממוזערת בצורה הכי משמעותית היא השגיאה הנפחית הכוללת, ואילו השגיאה המתקבלת ביד ימין קרובה מאוד לשגיאה ביד שמאל – כלומר יש כאן סימטריה.

ב. תוספת הפסד נפח ביד ימין :



גרף 5

ובאופן דומה – חישוב הפסד נפח על יד ימין מביא למזעור של השגיאה הנפחית ביד ימין, כאשר השגיאות באיברי הגוף האחרים יותר גבוהות.

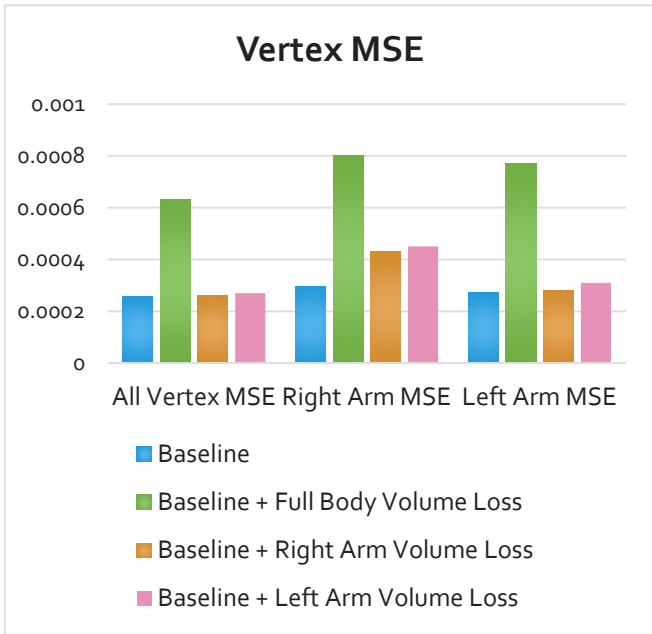
התוצאות הסופיות של שלב ה-Test רוכזו בטבלה הבאה, שמציגה מדדים של עיוות נפחי ושל MSE :

Experiment/metric	Baseline	Baseline + Full Body Volume Loss	Baseline + Right Arm Volume Loss	Baseline + Left Arm Volume Loss
All Vertex MSE	0.0002581	0.0006321	0.000261	0.00027
Right Arm MSE	0.0002973	0.0008013	0.0004295	0.00045
Left Arm MSE	0.0002709	0.000772	0.0002793	0.00031
Full Body Volume Deformation [%]	9.746	1.005	4.827	5.517
Right Arm Volume Deformation [%]	15.258	1.958	1.916	12.765
Left Arm Volume Deformation [%]	15.138	1.967	11.036	1.585

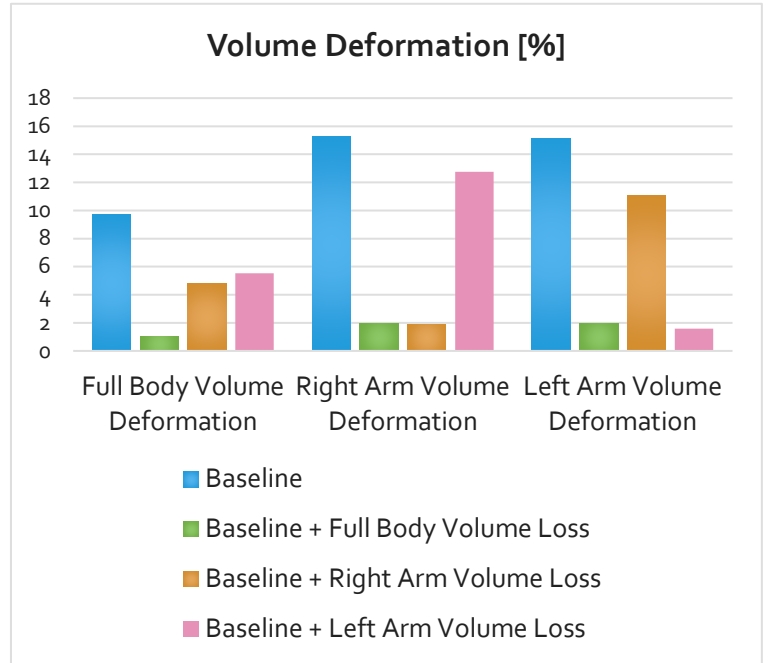
טבלה 2

בתוצאות המסכמות האלה ניכרת ההתאמה בין האיבר אליו התייחסנו בפונקציית ההפסד – לבין השגיאה המתקבלת. כמו כן, ניתן לשים לב לאותו trade off בין עיוות נפחי לבין מדד MSE. למשל, אם נתבונן בעמודה של Baseline + Right Arm Volume Loss לעומת Baseline, נבחין שהשיפור ב- Right Arm Volume Deformation מתקבל על חשבון פגיעה והגדלה של Right Arm MSE, בעוד שאין כמעט הבדל במדד ה- All Vertex MSE – והוא אדיש לאופטימיזציה המתבצעת על איבר בודד.

ניתן לראות את היחסים השונים בין מדדי השגיאה בכל ניסוי גם בגרפים הבאים :



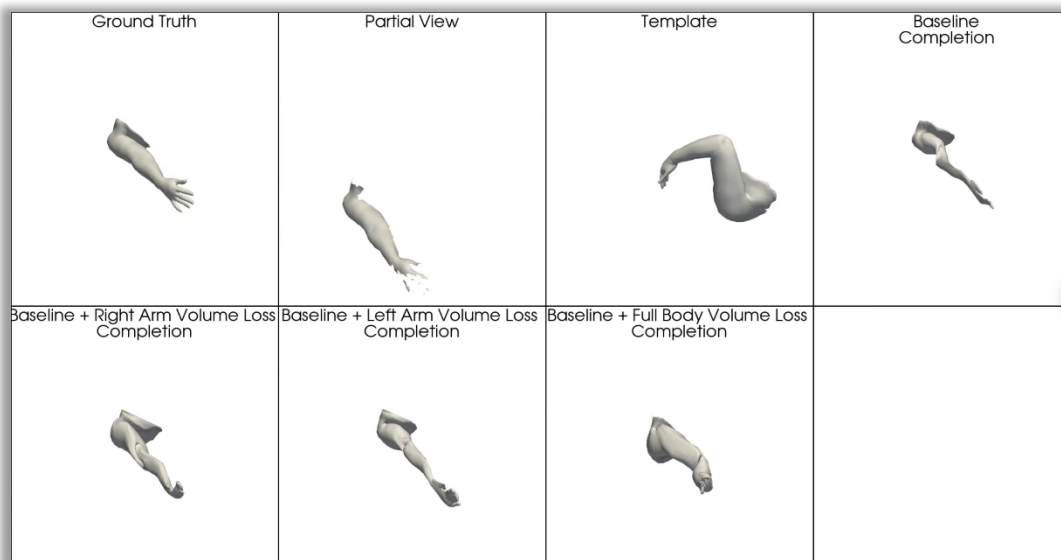
גרף 7



גרף 6

בהמשך, על מנת לנתח את התוצאות בצורה איכותית וויזואלית יותר, בחרנו מספר דוגמאות והעמקנו את הניתוח של התוצאות המתקבלות :

▪ דוגמא של יד ימין :



איור 12

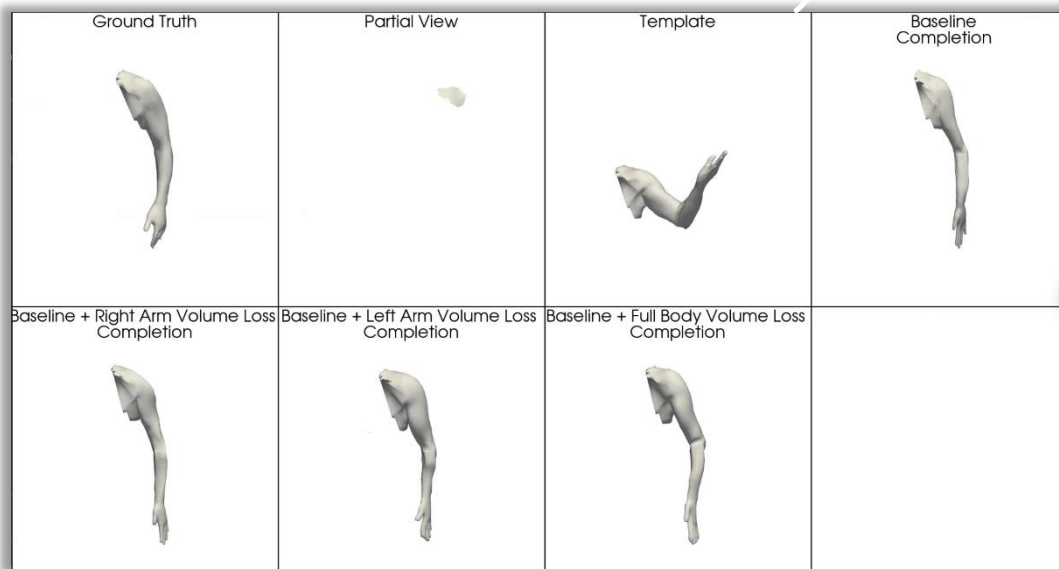
	Baseline	Baseline +Right Arm Volume Loss	Baseline +Left Arm Volume Loss	Baseline +Full Body Volume Loss
GT-comp Vertex MSE	0.000388	0.000774	0.000543	0.002692
Completion Volume	0.002719	0.005396	0.003469	0.005667
Volume Deformation [%]	47.7818	3.6182	33.3789	8.8304

טבלה 3

ניתן להבחין בתוצאות הנומריות שקיבלנו מה שציפינו – התוצאה הכי טובה במונח של אובדן נפח התקבלה כאשר הענשנו את המודל על טעויות ביד ימין. לאחר מכן, במקום השני, כאשר מענישים על טעויות נפח, ובשלישי את העונש רק על יד שמאל.

ניתן להבין מכך, שכאשר המודל מנסה לשפר את יד ימין בלבד, הוא מסוגל "להקריב" נכונות בשאר האיברים (ניתן לראות זאת בדוגמה של גוף מלא בהמשך), ולשאוף כמה שיותר לדייק בנפח של יד ימין. כל שינוי בפונקציית ההפסד, שכלל רכיב חדש, גרם לפגיעה כלשהי ב-Vertex MSE, וניתן להבין שמטרת האופטימיזציה המסובכת יותר הביאה אותנו לנקודה פחות אופטימלית עבור פונקציית ההפסד הזאת.

■ דוגמה של יד שמאל:



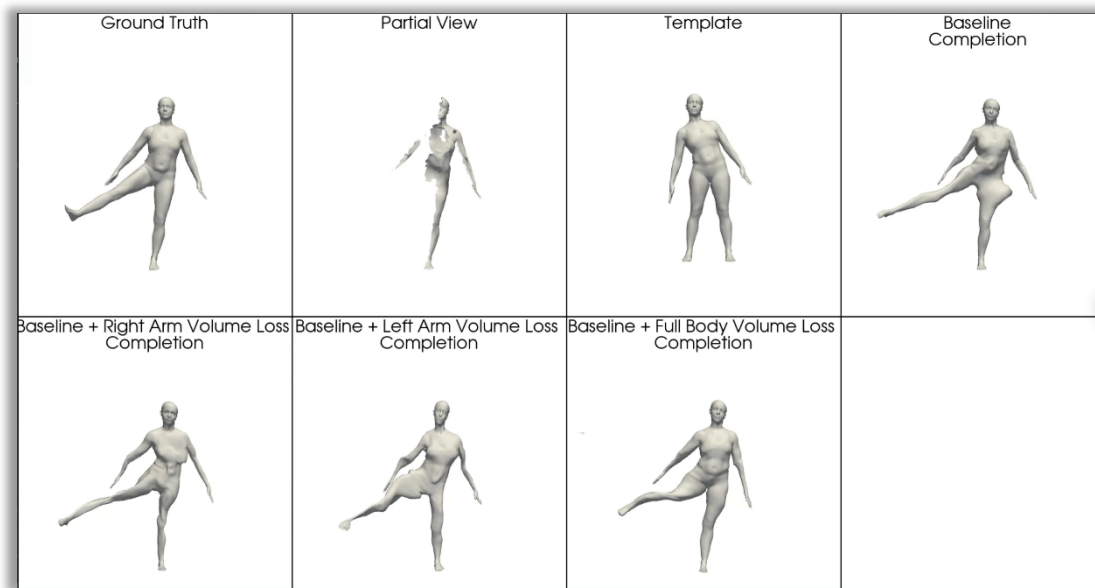
איור 13

	Baseline	Baseline + Right Arm Volume Loss	Baseline + Left Arm Volume Loss	Baseline + Full Body Volume Loss
GT-comp Vertex MSE	0.000924	0.000474	0.000546	0.000369
Completion Volume	0.003650	0.004096	0.006329	0.006094
Volume Deformation [%]	42.4609	35.4364	2.3926	3.9412

טבלה 4

קיבלנו תוצאה מאוד דומה לתוצאה שקיבלנו עבור יד ימין, וכך הדגמנו שהשיפור משפיע בצורה דומה מאוד אם מענישים רק על יד ימין, רק על יד שמאל, או כל הגוף.

דוגמה של גוף מלא:



איור 14

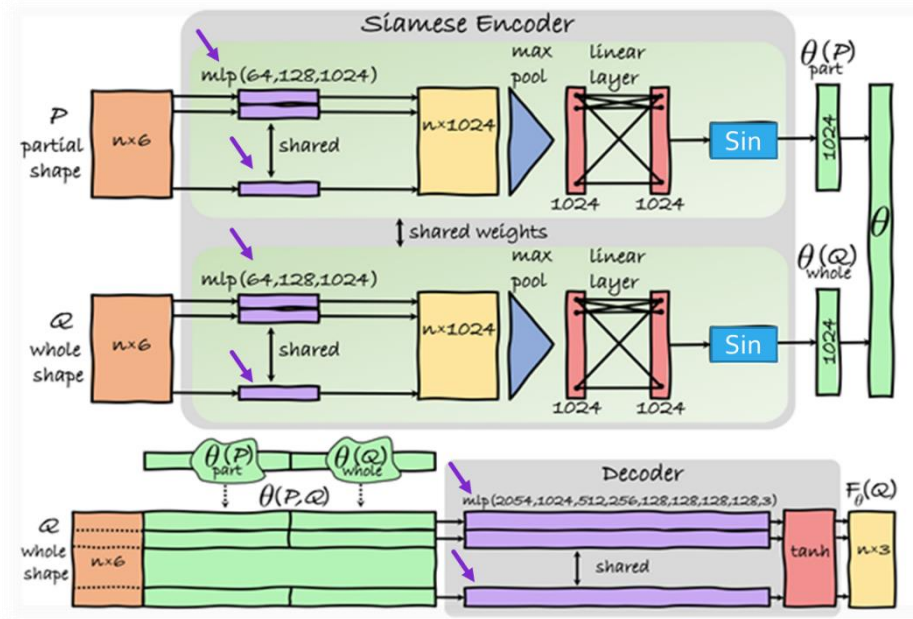
	Baseline	Baseline + Right Arm Volume Loss	Baseline + Left Arm Volume Loss	Baseline + Full Body Volume Loss
GT-comp Vertex MSE	0.001797	0.001568	0.001746	0.001753
Completion Volume	0.058038	0.051763	0.053462	0.066683
Volume Deformation [%]	14.8224	24.0387	21.5466	2.1455

טבלה 5

בדוגמה הזאת, אנו רואים דוגמה לכך שהענשה על סטייה מנפח כלל הגוף, מביאה תוצאות שנראות טוב יותר, מאשר שאר האפשרויות - להעניש על כל יד בנפרד גרר "פיצוי" או קומפנסציה - בריעת האופטימיזציה, נקודת האופטימום שהמודל נחת עליה ראתה לנכון לתקן את היד על חשבון שאר חלקי הגוף- דבר שמוביל לתוצאה ויזואלית לא איכותית מדי.

5.2 שינויים ארכיטקטוניים – החלפת פונקציות האקטיבציה בפונקצי סינוס :

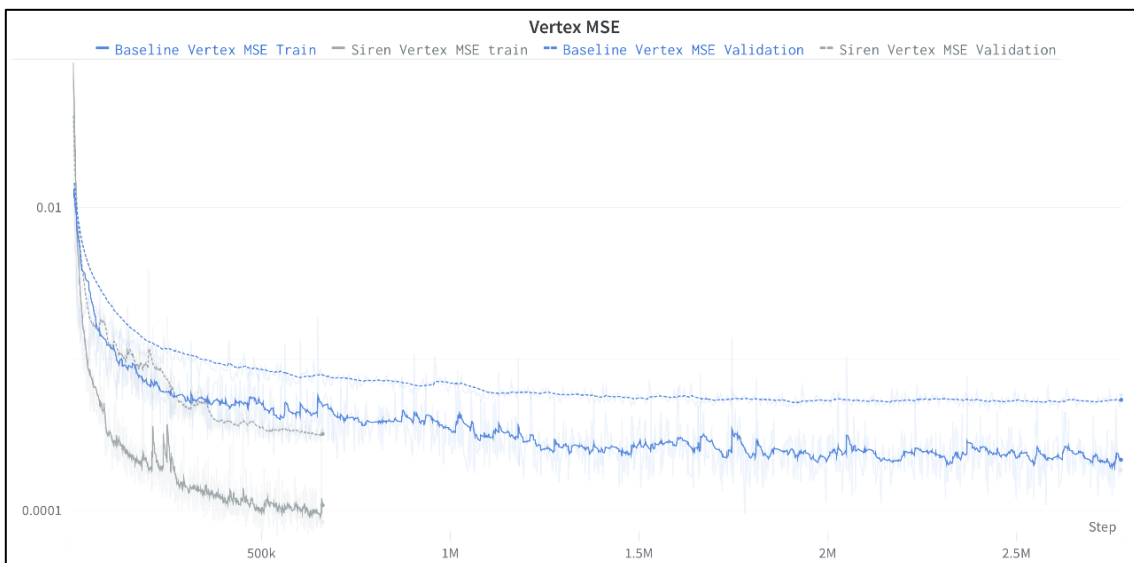
בניסוי זה, כאמור, החלפנו את פונקציות האקטיבציה ברשת בפונקציות סינוס, בהתאם למאמר אודות ארכיטקטורה מבוססת Siren. מבנה הרשת החדשה מתואר באופן הבא :



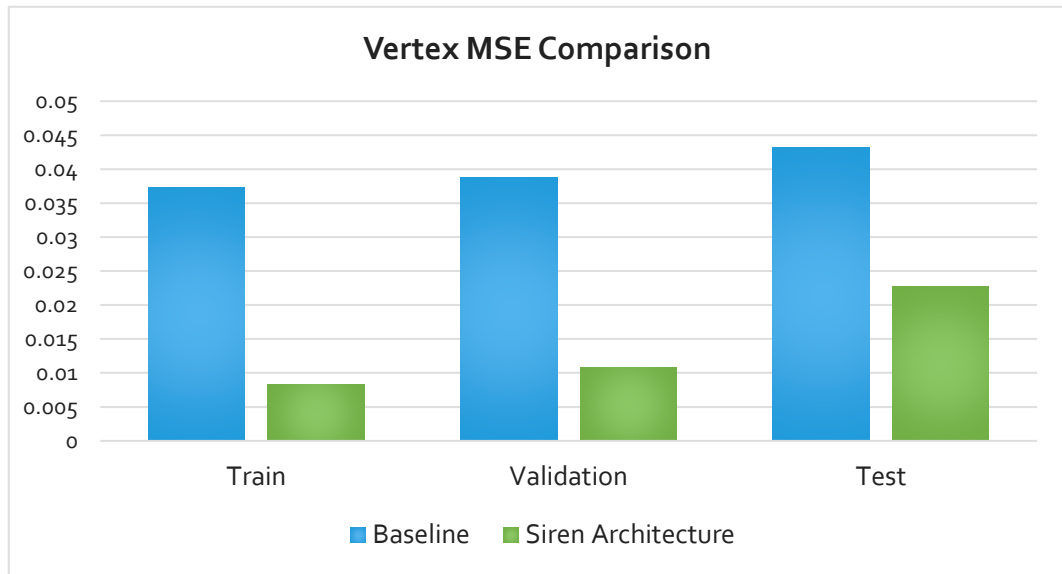
איור 15

*פונקציות האקטיבציה מוחלפות גם בתוך שכבות ה- MLP ברכיבי ה-Encoder וה-Decoder, כפי שמסומן בחיצים באיור.

במקרה זה, ביצענו ניסוי השוואתי פשוט – האם במובן Vertex MSE, שינוי פונקציות האקטיבציה לסינוסים מוביל לתוצאות טובות יותר במובן MSE. ניתן לראות שהניסוי עבר בהצלחה, והשיפור הוביל לתוצאות בעלות רזולוציה גבוהה יותר, ואיכותית יותר, בזמן התכנסות מהיר יותר, ולערך יותר אופטימלי מאשר הארכיטקטורה הבסיסית. השיפור אליו הצלחנו להגיע במימוש זה, ניכר בעיקר במדד ה-MSE של הצמתים :



גרף 8



גרף 9

ניתן לראות שהשיפור ב-vertex MSE מאוד משמעותי. הארכיטקטורה מצליחה ללמוד טוב יותר את צורת הדמויות, וברזולוציה יותר גבוהה מהארכיטקטורה הבסיסית בה התחלנו. ניתן לראות זאת גם בתוצאות ויזואליות. ניתן לראות בגרף האימון שההתכנסות הייתה מהירה בהרבה (כמו שרואים שקורה במאמר המקורי של siren).

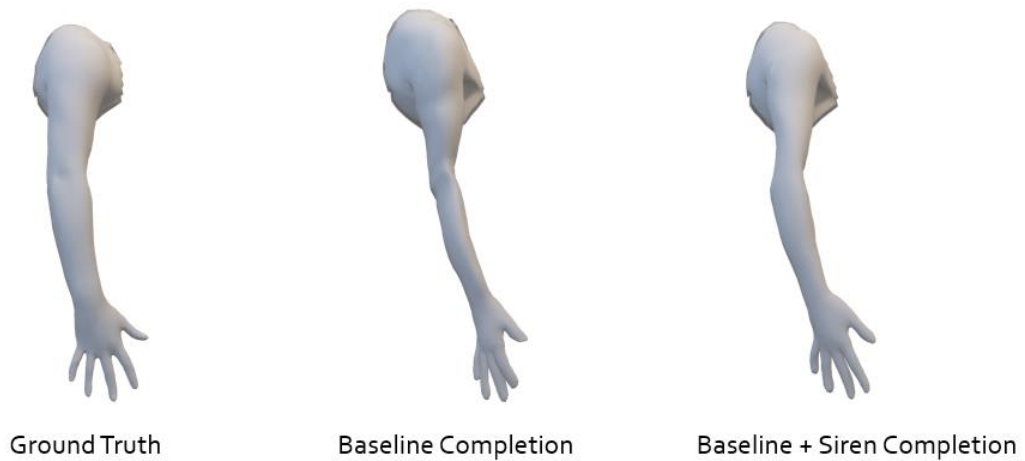
גם במימוש זה, בחרנו מספר איברים לניתוח מעמיק יותר, מאחר ובידיים (כמו שנסביר בהמשך) ניתן לראות את ההבדלים בתדרים גבוהים:



איור 16

	Baseline	Baseline +Siren Architecture
GT-comp Vertex MSE	0.0009158	0.0005248

טבלה 6



איור 17

	Baseline	Baseline +Siren Architecture
GT-comp Vertex MSE	0.0009404	0.0004877

טבלה 7

בשתי דוגמאות אלה, ניתן לראות את אותה תופעה מתרחשת – פונקציית האקטיבציה החדשה מובילה לתוצאות חדות ו"יפות" יותר – יותר קרובות לGround Truth, ויותר מתאימות לתוצאה רצויה. ניתן לראות את השיפור במיוחד באזור האצבעות – אצבעות בסריקות של DFAUST הן האזור בעל תדרים הכי גבוהים שיש במערכת – המעבר בין קצה הבוהן, המפרקים, הוא מאוד עדין, וניתן לראות שבבסיס Baseline יש אובדן משמעותי בפרטים באזורים אלה, לעומת siren שבה ניתן לראות את צורת היד בצורה חדה.

5.3 ניסויים נוספים – ארכיטקטורת LSTM

ניסינו, כאמור, להשתמש גם ב-Decoder המותאם לארכיטקטורת LSTM, לשיפור החלקות של השחזורים.

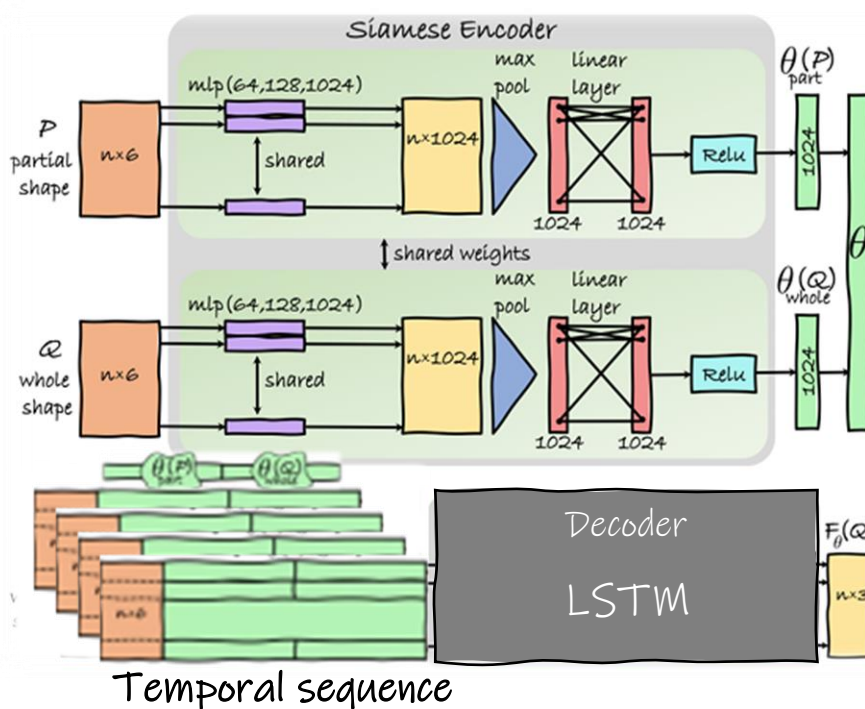
LSTM (Long short-term memory) היא רשת נוירונים מלאכותית, המיועדת לשימוש בתחומי הבינה המלאכותית והלמידה העמוקה.

בניגוד לרשתות זרימה קדימה, רשת LSTM מכילה קשרי משב.

רשת כזו יכולה לעבד מלבד נקודות מידע בודדות, גם רצפים שלמים של מידע.

רשתות אלה מתאימות לסיווג, עיבוד וחיזוי בהתבסס על מידע טמפורלי.

הבסיס של השינוי שביצענו הוא בהחלפת המפענח של הרשת המקורית במפענח LSTM:



איור 18

על מנת לממש את הרשת באופן הזה, נדרשנו לממש מנגנון טעינה חדש. הרשת בארכיטקטורה הבסיסית שלה לא קיבלה מידע בהקשר זמני. כלומר, ניתן היה לקחת כל סריקה בנפרד, באופן רנדומלי, להכניס מידע זה לרשת ולאמן אותה. כאשר מוסיפים דרישות זמניות לרשת, יש צורך במנגנון חדש.

מימשנו מנגנון זה באופן הבא (נתאר זאת בצורת אלגוריתם):

קלט:

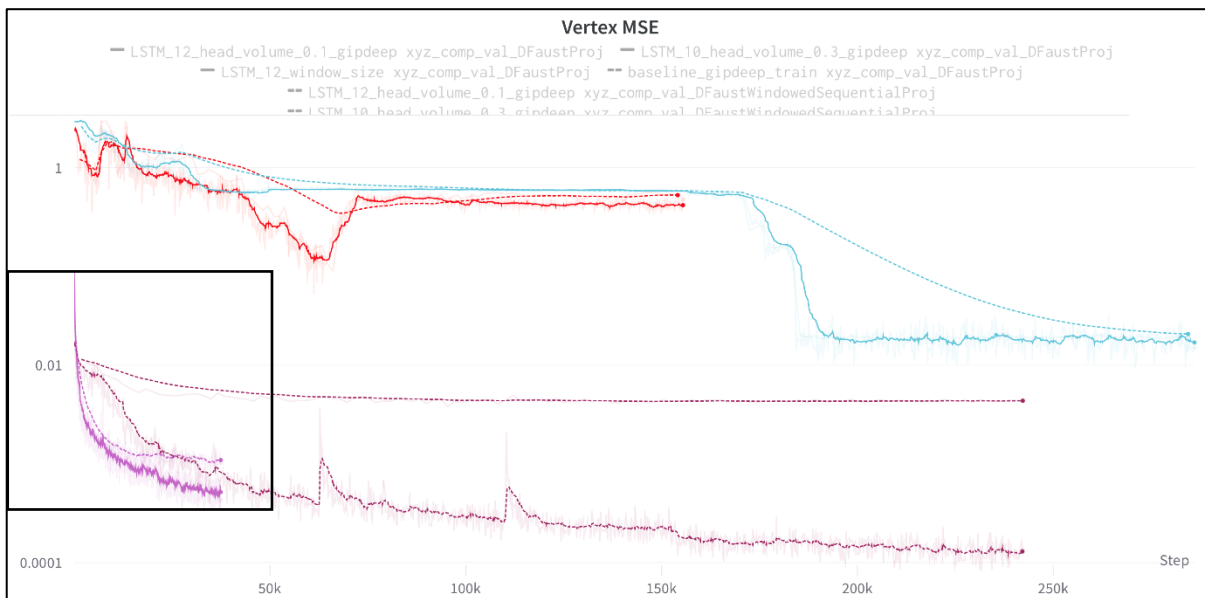
1. אינדקס סריקה רנדומית כלשהי
2. גודל חלון רצוי
3. גודל קפיצה זמנית בין הסריקות, שיכללו בחלון

פלט: אצווה של סריקות, כאשר כל סריקה מייצרת אצווה שלמה בגודל החלון

אלגוריתם:

1. רוץ אחורה, החל מהאינדקס של הסריקה הראשונית, בגודל קפיצה שהתקבל בקלט
2. כל עוד לא הגענו לאינדקס שלילי (כלומר, במהלך הזמן, לא הגענו לסריקה בעלת אינדקס בתוך האינמציה, הקטן מאפס), כלול את הסריקה, והקטן את האינדקס שנית
3. בהינתן והגענו לערך שלילי, כלול את הסריקה באינדקס ה-0.
4. חזר את האצווה שנוצרה

לאחר מכן, לאחר שמימשנו מנגנון טעינה חדש, החלפנו את המפענח ברשת למפענח חדש, ביצענו מספר ניסויים שונים.



גרף 10

מקרא:

בקו רציף – Train, בקו מקווקו – Validation
תכלת – LSTM עם גודל חלון 12 וערך קפיצה של 8, וערך 0.1 למקדם בפונקציית ההפסד על נפח הראש
ורוד – LSTM עם גודל חלון של 12 וערך קפיצה 8
אדום – LSTM עם גודל חלון 10 וערך קפיצה של 8, וערך 0.3 למקדם בפונקציית ההפסד על נפח הראש
שני הגרפים במסגרת – שייכים לריצת ה-baseline

הניסויים על נפח הראש בוצעו מכיוון שהפנים הם החלק הכי מעוות בהשלמות הבסיסיות שהתקבלו במימוש זה, וניסינו להילחם בעיוות זה בפונקציית המטרה.
*יש לציין שביצענו ניסויים רבים נוספים, ובחרנו לשתף חלק מהניסויים באופן מדגמי, מאחר שהתוצאות לא היו משביעות רצון בכל מקרה.

רכיב מהירות לפונקציית ההפסד:

הוספנו רכיב ממשקל לפונקציית ההפסד, שמחשבת את "מהירות" הצמתים בענן הנקודות, שמטרתו היא לנסות להבטיח חלקות מיטבית יותר לתוצרים הסופיים אותם יצרנו.

הוא מוגדר כך :

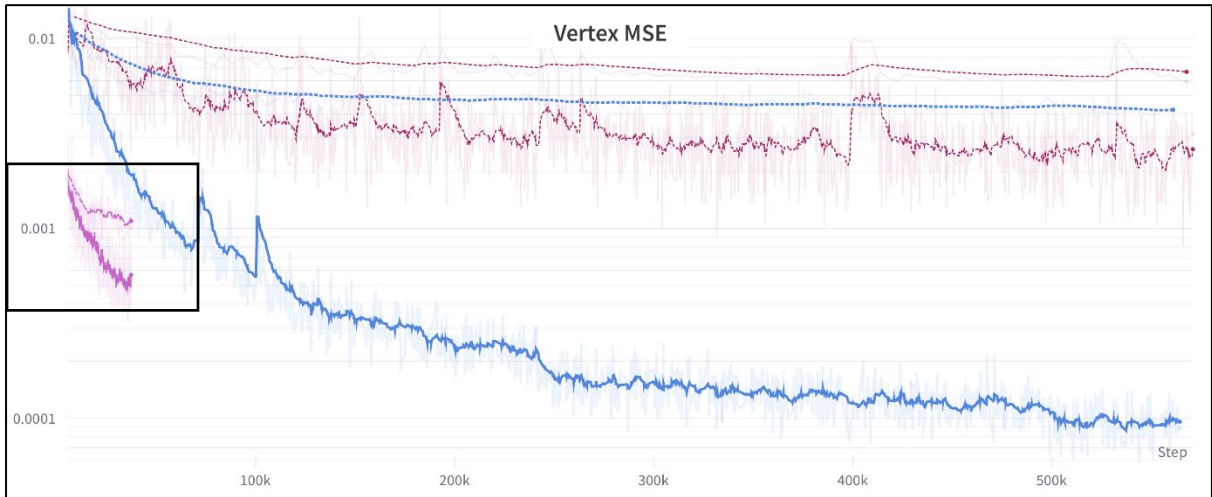
עבור השלמה X , ודוגמת Y Ground Truth, נגדיר :

$$V_{Y_t} = Y_{t+1} - Y_t$$

$$V_{X_t} = X_{t+1} - X_t$$

$$L_{Velocity Loss} = |V_{Y_t} - V_{X_t}|_2^2$$

ראינו שהתוצאות המתקבלות לא מראות על התכנסות :



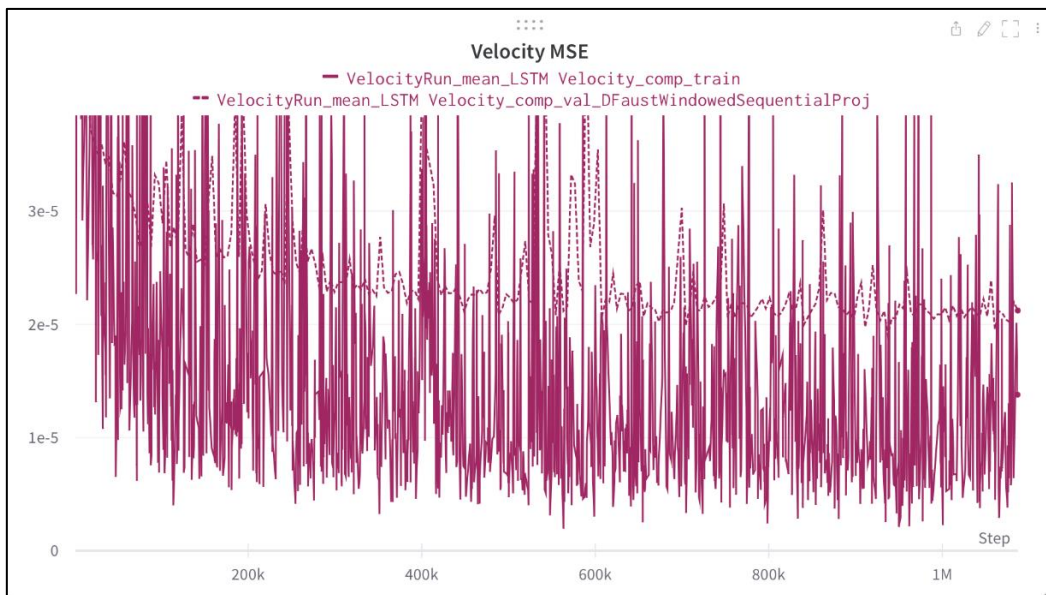
גרף 11

מקרא:

בקו רציף – Train, בקו מקווקוו – Validation

כחול – LSTM עם גודל חלון 12, ורוד – LSTM בתוספת רכיב מהירות

שני הגרפים במסגרת – שייכים לריצת ה-baseline



גרף 12

ניתן לראות מגרף האימון, שהשיטה לא הצליחה להתכנס מבחינת יציבות האימון, ואף התכנסה לתוצאה לא טובה ויזואלית כלל (לא חלקה יותר במובן טמפורלי, ומעוותת גם במנח סטטי).
ניתן לראות זאת בדוגמאות:

50009_hips



50026_shake_arms

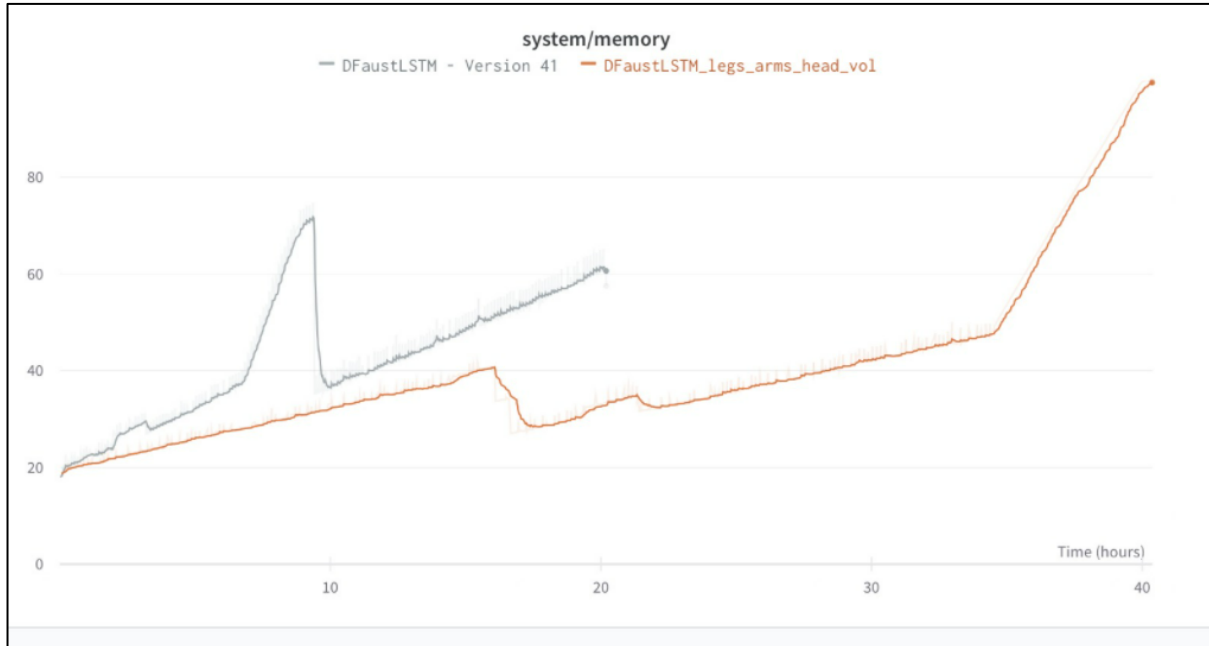


איור 19

השיטה נכשלה בלספק לנו תוצאות טובות, כפי שניתן לראות בגרף האימון, ובנוסף, ויזואלית, הרשת לא הצליחה ללמוד כיצד להשלים בצורה איכותית.

6 אתגרים ופתרונות

א. דליפת זיכרון:



גרף 13

גילינו במהלך הפרויקט כי ישנה אפשרות לדליפות זיכרון ב-GPU ב-pytorch, והבעיה נוכחת כאשר השתמשנו ב-LSTM Decoder.

הצלחנו לצמצם את הבעיה, על ידי שינוי ה-loader שיעבוד בצורה יותר נכונה, ותיקון שגיאות קטנות במימוש. גם לאחר התיקונים, עדיין חווינו דליפת זכרון. מאחר וה-LSTM בנוסף לא הראה תוצאות איכותיות מספיק, החלטנו לא להתעמק בתיקון מעבר לכך.

ב. שיטות logging ואיטיות שרתים:

התחלנו לעבוד בשלב מוקדם בפרויקט עם Weights and Biases. לאחר זמן מה, הגענו למצב שתיעוד התוצאות בשרתים האינטרנטיים האט בצורה דרמטית את ריצת האימון שלנו (בסדר גודל). לאחר אבחון וביצוע תהליך debug יחד עם התמיכה הטכנית של מפתחי הכלי, הבנו שהבעיה הייתה במספר הגרסאות שיצרנו, של טבלאות או של ערכים שתיעדנו. תיקנו בעיה זאת, וצמצמנו באופן כללי את היקף התוצאות שתיעדנו בשרתים, וכך הצלחנו להתמודד עם הבעיה.

7 סיכום ומסקנות

הצלחנו לספק שיפורים משמעותיים ברזולוציה של ההשלמות, דבר שנוגע ללב המטרה במאמר המקורי, ומשפר את איכות ההשלמות המדויקות בצורה אלגנטית ופשוטה. בנוסף, הצגנו שיטה נוספת להילחם בעיוותים המתקבלים במאמר המקורי, כמו עיוות הנפח. תיקוני עיוות הנפח, הם קונפיגורציות ונשלטים, כלומר, ניתן לקבוע כמה עיוות אנו מוכנים לאפשר בכל איבר בנפרד, או שאנו מעוניינים לנסות לתקן את כולו. שיפורים אלו הם אורתוגונליים, ואינם תלויים אחד בשני. הגענו למסקנה כי לפחות בארכיטקטורה שבחרנו לממש, השיפורים הזמניים אותם ניסינו לעשות נכשלו, ולא הצליחו להביא לתוצאות הרצויות שניסינו להגיע אליהם. במידה והיינו מתחילים את הפרויקט היום, כנראה היינו מנסים להשתמש בשיטות יותר חדשות ויציבות לאימון, לדוגמה שיטת transformers. תחת המימוש בו בחרנו, פונקציות הפסד זמניות לא תפקדו כראוי, ולא היו יציבות לאימון, וייתכן ויעבדו טוב יותר בהינתן והיינו משתמשים בארכיטקטורה יותר יציבה לאימון.

חוינו קשיים במהלך העבודה על הפרויקט עם שיטות שנכשלו, וביצוע מחקר, עולם חדש ולא מוכר עבורנו. למדנו מכך שבמהלך העבודה, חשוב לבדוק ולוודא כל חלק בנפרד, ולערוך ניסויים מתאימים לטובת וולידציה מלאה ואיכותית. הפרויקט עזר לנו ללמוד כיצד לבצע זאת יותר טוב בהמשך.

השתמשנו בכלים חדשניים ולמדנו איך נראית עבודת המחקר, ונשתמש בכלים אלה בהמשך עתידנו המקצועי והאקדמי. למדנו לנתח תוצאות, גם בצורה נומרית וגם ויזואלית, וכמה חשובה יכולת ההדגמה המספרית הזאת, כדי להוכיח נכונות גם לעצמנו, וגם להצגת עבודתנו לאחרים.

8 תודות

נרצה להודות לעומר בן-חיון, על עזרתו הרבה בפיתוח התשתית והאדיבות בשיתוף הידע שלו.

9.1 מרכז של הדמויות כלפי מרכז המסה

רצינו לבחון את ההשפעה של ביצוע טרנסלציה של מרכזי הכובד של ענני הנקודות סביב ראשית הצירים $(0,0,0)$ ובכך להבטיח שמדידות המטריקות יתבצעו בהתאם למיקומים הנכונים. ניתן לגשת לצפות בתוצאות בקישור:

<https://www.youtube.com/watch?v=SLBte-Bh30w>

תוספת זו מאפשרת לנו לבטל offset אנכי המתקבל בחלק מההשלמות המקוריות, וגם להבטיח תנועה חלקה יותר. כחלק מהניסוי, החלטנו לבצע את המרכז, כלפי מרכז המסה של פלג הגוף העליון- כלומר ביצענו חישוב מרכז מסה על כל הצמתים המרכיבות את פלג הגוף העליון של הדמויות התלת-מימדיות. כצפוי, התוצאות היו חלקות בהרבה, אך נמנעו מתנועה אנכית משמעותית. נראה מכך שסכמת המרכז המקורית לא הייתה משביעת רצון, כמו שהסכמה החדשה, על יתרוניתה, לא מספקת את כלל הדרישות שהיינו מעוניינים להשיג – ביצוע מרכז של כל הדמויות כלפי אותה מערכת צירים, תוך כדי אפשר תנועה אנכית איכותית.

10 שרטוטים ואיורים

4	איור 1
4	איור 2
5	איור 3
6	איור 4
6	איור 5
8	איור 6
9	איור 7
10	איור 8
10	איור 9
11	איור 10
12	איור 11
17	איור 12
18	איור 13
19	איור 14
20	איור 15
21	איור 16
22	איור 17
23	איור 18
26	איור 19

11 גרפים

14	גרף 1
14	גרף 2
15	גרף 3
15	גרף 4
16	גרף 5
17	גרף 6
17	גרף 7
20	גרף 8
21	גרף 9
24	גרף 10
25	גרף 11
25	גרף 12
27	גרף 13

12 טבלאות

9	טבלה 1
16	טבלה 2
17	טבלה 3
18	טבלה 4
19	טבלה 5
22	טבלה 6
21	טבלה 7

¹ <https://brickisland.net/DDGSpring2020/>

² http://groups.csail.mit.edu/gdpgroup/6838_spring_2019.html

³ <https://www.meshlab.net/>

⁴ Qi, Charles R., Hao Su, Kaichun Mo, and Leonidas J. Guibas. "Pointnet: Deep learning on point sets for 3d classification and segmentation." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652-660. 2017.

⁵ Halimi, Oshri, Ido Imanuel, Or Litany, Giovanni Trappolini, Emanuele Rodolà, Leonidas Guibas, and Ron Kimmel. "Towards precise completion of deformable shapes." In *European Conference on Computer Vision*, pp. 359-377. Springer, Cham, 2020.

⁶ Sitzmann, Vincent, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. "Implicit neural representations with periodic activation functions." *Advances in Neural Information Processing Systems* 33 (2020): 7462-7473.