

1. על מנת לחקות CNN באמצעות BERT נרצה להזין את הרשת עם חלקים (פאצ'ים) של התמונה, בצירוף מידע על המיקום של הפאצ'. זאת מכיוון שאנו רוצים ש-BERT יבדוק את הקשר בין איזורים קרובים ולא רחוקים, בדומה למה שה-CNN עושה בסדר גודל של ה-kernel שנקבע מראש. באופן כזה תשמר הפעולה של BERT על הקלט המקורי שהוא חלקים מהטקסט (משפטים) ומידע על המיקום שלהם, מה שנותן לרשת את היכולת ללמוד את ההקשר בין המשפטים, ובמקרה שלנו את הפיצורים של התמונה. בחיקוי זה נשתמש בכל פאצ' כיחידה הבסיסית של מנגנון ה-Self Attention. מכיוון שפעולת Self Attention בודקת את הקשר בין כל זוג רכיבים אטומיים בדאטה, היא אינווריאנטית למיקום כמו CNN ולכן עומדת בדרישה זו לחיקוי CNN. בנוגע לחיקוי MLP, אם אנחנו מזינים את BERT עם כל התמונה כולה (ולא פאצ'ים שלה), אז הרכיב האטומי של התמונה יהיה פיקסל, ובכך תתבצע השוואה בין כל שני פיקסלים בתמונה, בדומה למה ששכבת FC או MLP היו מבצעים.

2.

Regular GAN

הדאטה-סט שהכי מתאים למודל זה הוא הדאטה-סט הראשון (הדאטה-סט בו ישנם זוגות ברורים של תמונות מ-A ומ-B). בבעיה זו אנו רוצים למצוא מיפוי טוב מ-A ל-B. משום שמרחב החיפוש הוא מאד גדול, היינו רוצים לצמצם אותו ולמקד את הגנרטור. בדאטה-סט זה, הגנרטור צריך לבצע שינויים קטנים (סגנון) בקלט כדי להצליח למזער את ההבדל בין B ל-G(A). דאטה-סט זה עדיף על הדאטה-סט השני (הדאטה-סט בו המידע מחולק לזוגות אך בגסות) מכיוון שמרחב החיפוש בו מדויק יותר מאשר בשני. מודל זה לא מתאים לדאטה-סט השלישי (זה שלא מחולק לזוגות). דאטה-סט זה לא מתאים למודל של למידה מפוקחת- למעשה זה דאטה-סט נטול תגים. מכיוון שאין לנו דרך לדעת איך כל תמונה מהתחום אמורה להראות בטווח, עבור GAN רגיל מרחב החיפוש של פונקציית המיפוי שנפרש ע"י הדאטה-סט השלישי גדול מידי וללא אילוצים נוספים (שיש למשל ב-cycle gan) יהיה למודל קשה להצליח במשימה.

Cycle GAN

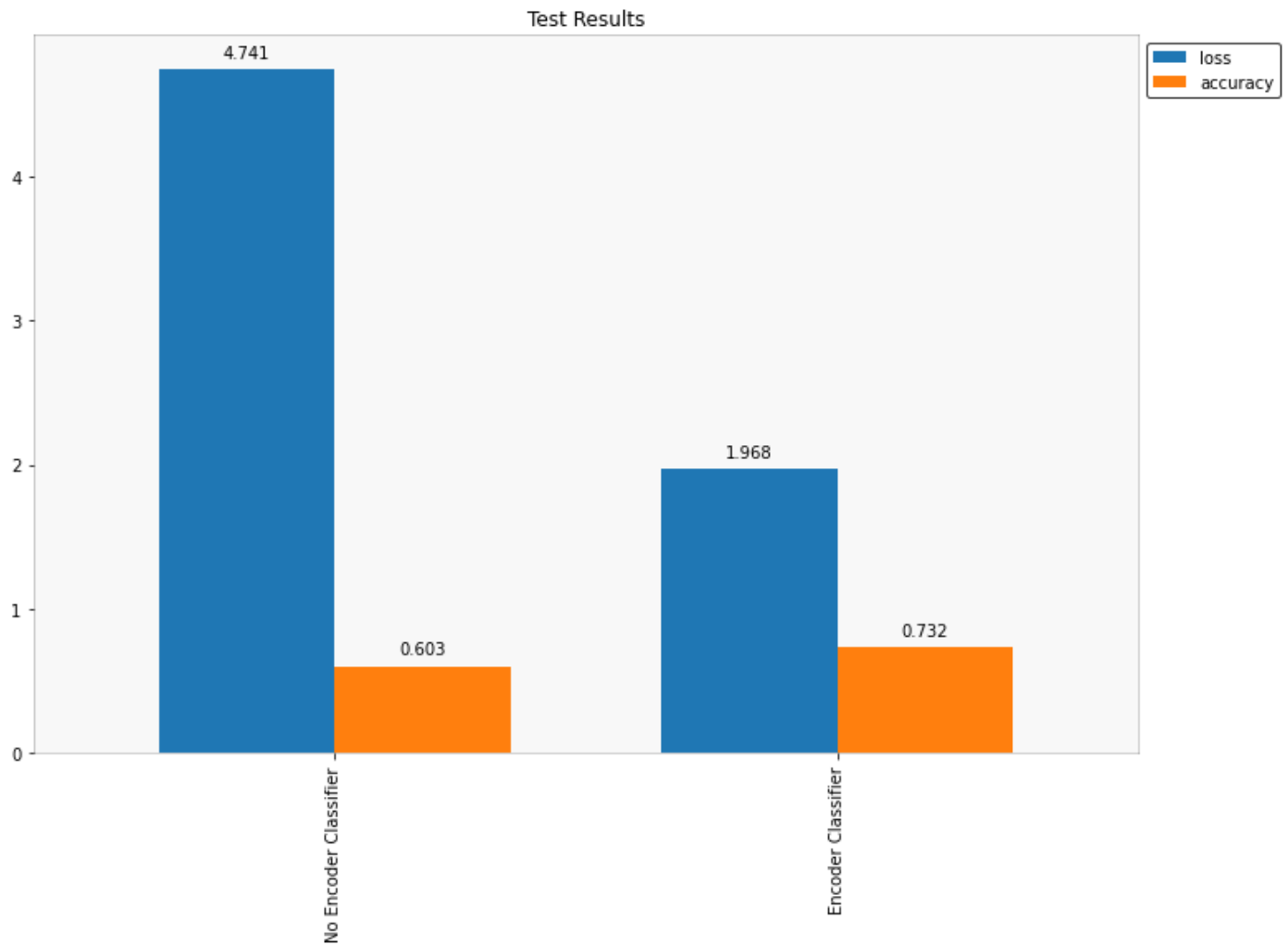
מודל זה הוא המודל היחיד שיעבוד יחסית טוב עם הדאטה-סט השלישי (Unpaired) מכיוון שהוא משתמש ב-cycle consistency loss. מודל זה מחפש שני מיפויים, $G_A: A \rightarrow B$, $G_B: B \rightarrow A$ כך ש- $G'_A = G_B$. השימוש ב-cycle consistency loss מאלץ את המודל לחפש את שני המיפויים כך שהמרחק בין $G_A(G_B)$ לבין העתקת הזהות Id קטן מאד. אילוץ זה עשוי לצמצם משמעותית את מרחב החיפוש ובכך לעזור לרשת להתמודד עם המטלה של unpaired dataset. אנו מצפים שניתן להשתמש במודל זה עם דאטה-סטים שהם paired מכיוון שעבודה איתם נותן צמצום גדול עוד יותר של מרחב החיפוש של פונקציית המטרה. החסרון של Paired הוא שהוא דורש כמות גדולה של עבודה אנושית לסידור הדאטה-סט, ולכן העובדה שGAN יכול להצליח במשימות על unpaired dataset מהווה יתרון משמעותי למודל זה.

Conditional GAN

מודל זה יתמודד היטב עם שני הדאטה-סטים הראשונים (paired dataset), מכיוון שמרחב החיפוש של פונקציית המיפוי שנפרש ע"י שני הסטים האלו הוא יחסית מצומצם. בנוסף, מודל זה יעבוד טוב יותר מה-GAN הרגיל מכיוון שהוא יכול לנצל את המצאות הלייבים כדי לבצע עליו התניה, כלומר להכניס כקלט את x,y יחד ובכך לתת לרשת אינפורמציה על איך להתגבר על חוסר הדיוק בכל צמד. מודל זה לא מתאים לדאטה-סט השלישי (זה שלא מחולק לזוגות) מאותם שיקולים שה-GAN הרגיל לא מתאים לכך.

פלוטים

סעיף 2

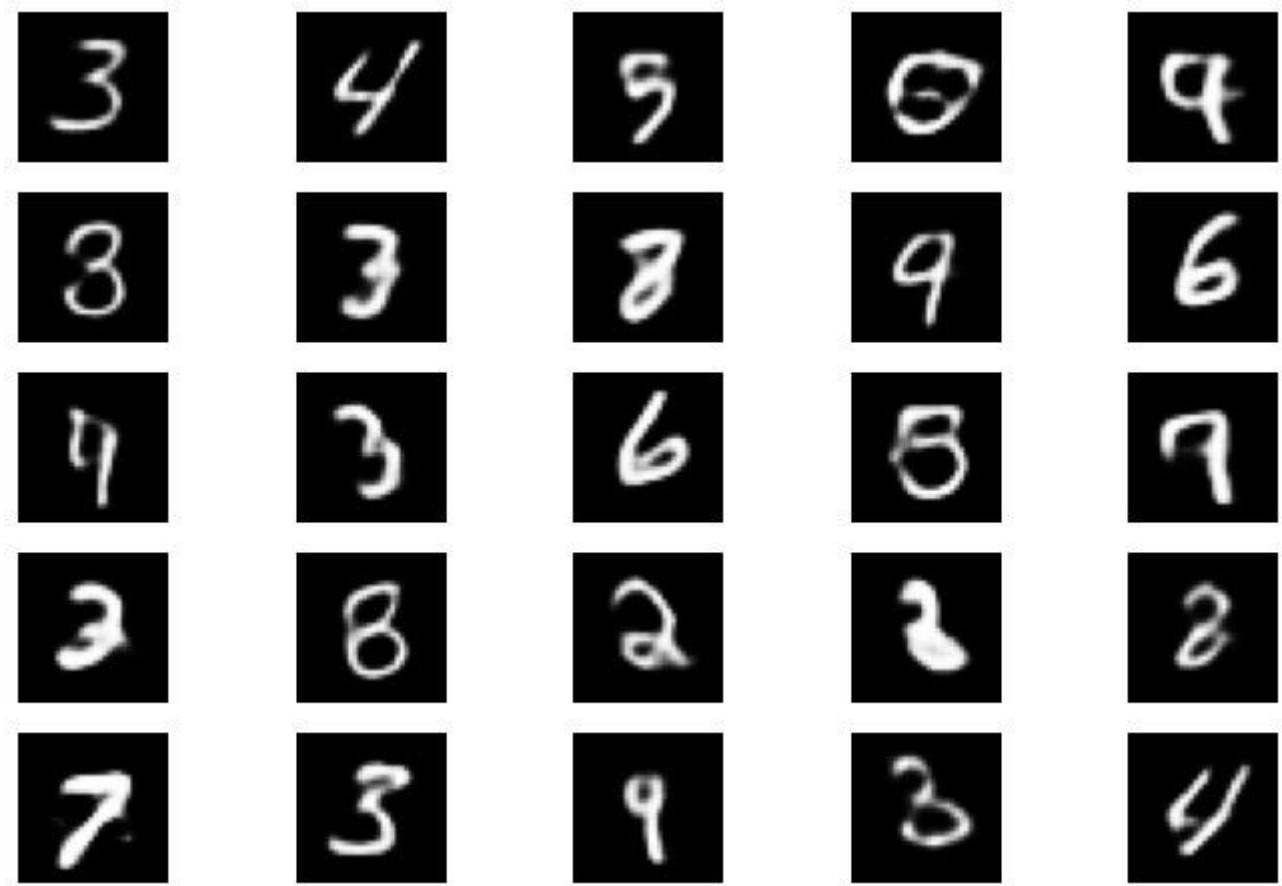


בארכיטקטורה זו השתמשנו ב-ELU ולא ב-ReLU, ובאופטימיזציה השתמשנו ב-rmsprop ולא ב-Adam, מכיוון שהם הביאו תוצאות טובות יותר.

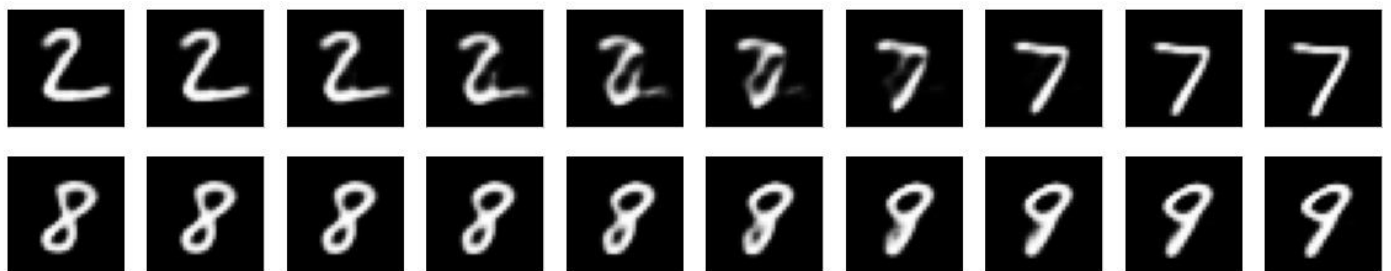
```
Model: "sequential_133"
```

| Layer (type) | Output Shape | Param # |
|-------------------------|--------------|---------|
| dense_455 (Dense) | (None, 64) | 4160 |
| dense_456 (Dense) | (None, 32) | 2080 |
| dense_457 (Dense) | (None, 10) | 330 |
| Total params: 6,570 | | |
| Trainable params: 6,570 | | |
| Non-trainable params: 0 | | |

סעיף 3
דוגמה לתוצאות של דגימה:



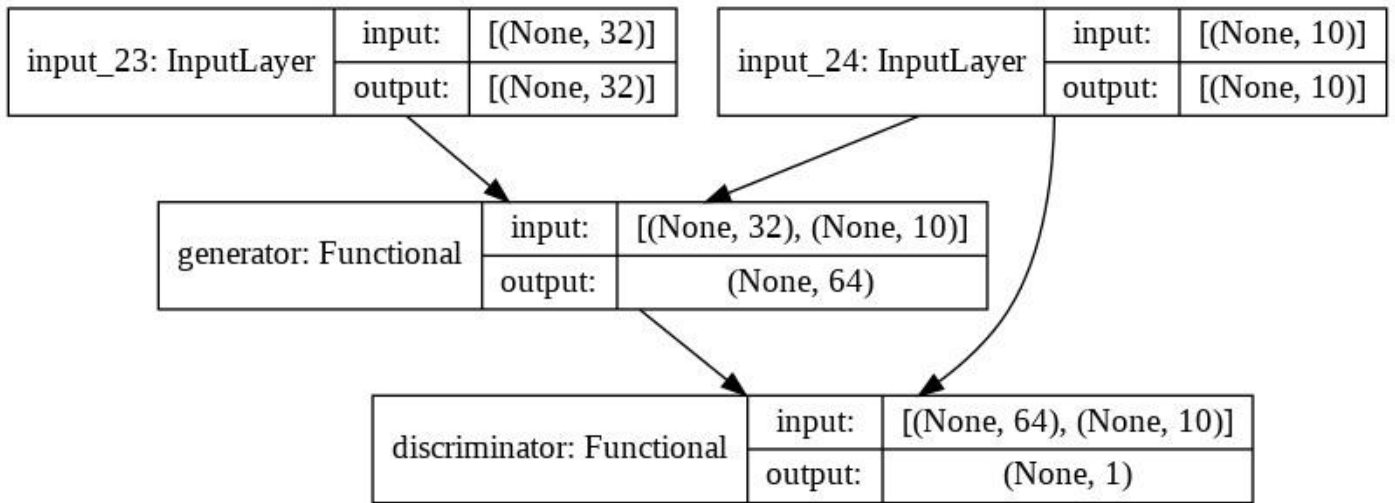
אינטרפולציה של ספרות (העליון- אינטרפולציה באמצעות האנקודר, והתחתון- אינטרפולציה באמצעות ה-GAN)



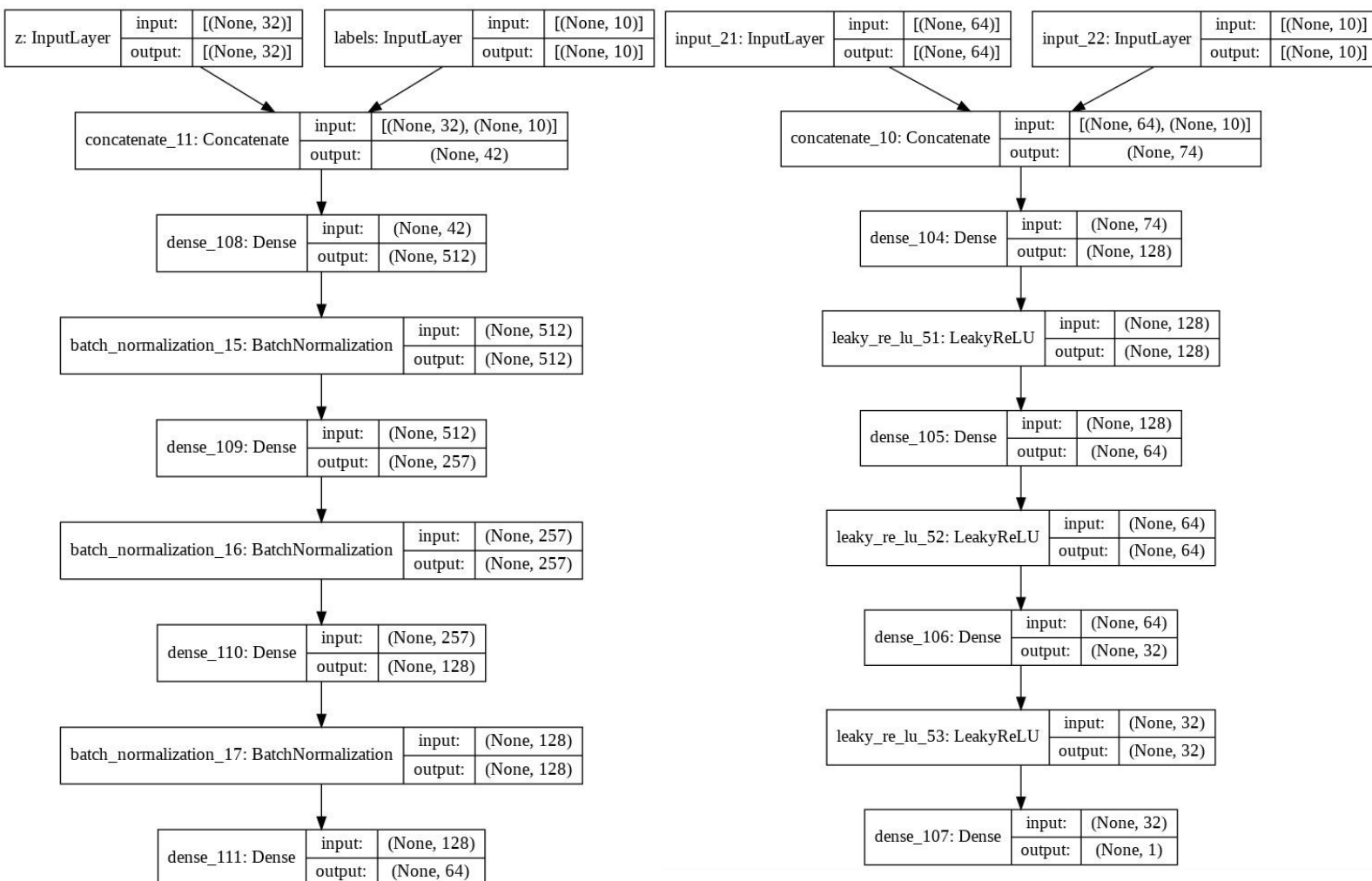
בקטע זה a (משתנה האינטרפולציה) קיבל עשרה ערכים מהקטע $[0,1]$. בצד שמאל הערכים הקרובים ל-0 ובצד ימין הערכים הקרובים ל-1.
ניתן לראות שבחלק התחתון (שיוצר ע"י GAN) ישנן תוצאות איכותיות יותר עבור ערכים שקרובים ל-0.5.
באנקודר, ערכים בסביבה זו לא נופלים במרחב ההתפלגות של המחלקה, לעומת ה-GAN, שעבורו ערכים בסביבה זו כן נמצאים בהתפלגות של המחלקה ולכן הוא נותן תוצאות הגיוניות יותר.

סעיף 4

בחרנו להשתמש ב-Conditional GAN, והוספנו את הלייבלים כאינפורמציה שמקודדת ב-one hot ונתונה גם לגנרטור וגם לדיסקרימיננטור. בתמונה הראשונה ניתן לראות את הארכיטקטורה של ה-GAN כולו, ובתמונות הבאות ניתן לראות את הפירוט של הארכיטקטורות של הגנרטור והדיסקרימיננטור.



הימני : דיסקרימיננטור, והשמאלי : גנרטור.



התוצאות :

