פתרון תרגיל מספר 2 ־ רשתות נוירונים בתמונות

שם: מיכאל גרינבאום, **ת.ז:** 211747639, **שם:** ניבי שנקר, **ת.ז:** 207227687

20 במאי 2021

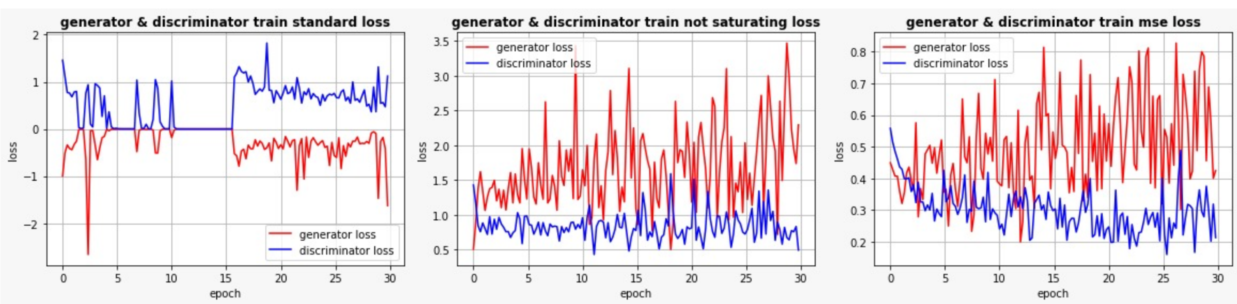
1. **חלק מעשי:**

(א) סעיף 1:

צ"ל: $saturation$

הוכחה:

נסתכל על הגרפים הבאים:

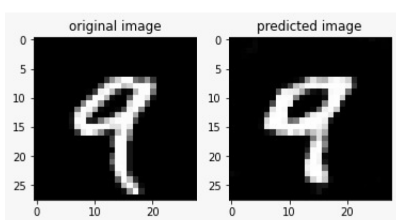


נשים לב שרק הראשון הוא $saturating$ וניתן לראות את זה על ידי הקושי של ללמוד בהתחלה. הוא בקושי יכול להשתפר וה־ $discriminator$ לומד הרבה יותר מהר ממנו.

צ"ל: שחזור תמונה

הוכחה:

סעיף 2:

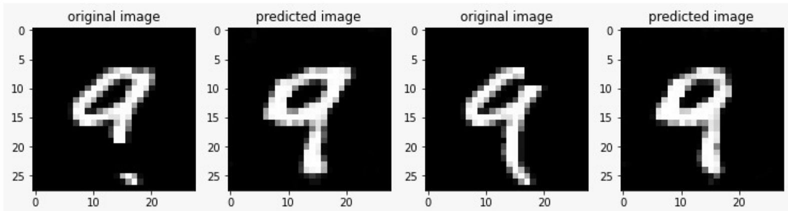


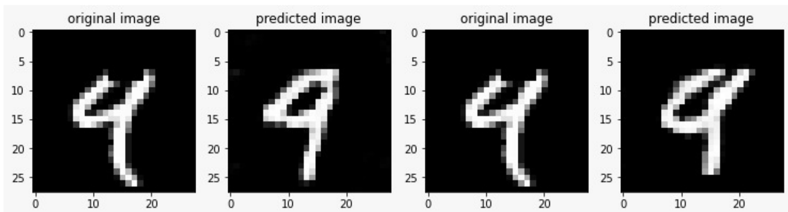
אפשר לראות שהרשת מצליחה לשחזר את ה־ 9 אבל לא במדויק, כנראה בגלל הגודל הקטן של הוקטור z שנדגם.

צ"ל: מחיקת חלקים בתמונה ושחזורה

הוכחה:

סעיף 3:

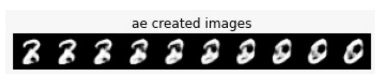




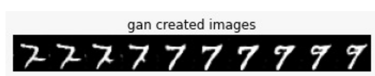
אפשר לראות שהוא מצליח לשחזר את העובדה שזה 9 על אף שאין זנב או חלק מהעיגול של ה־ 9 חסר.
רואים שקשה לו לשחזר ספרה כשהיא דומה לספרה אחרת (במקרה שלנו 9 דומה ל־ 4) ואז הוא משחזר פחות פרטים מהתמונה ואף קשה לדעת מהתמונה שנוצרה האם זה 4 או 9.

מ.ש.ל.א.☺

(ב) **צ"ל:** ללמד VAE ולעשות אינטרפולציה
הוכחה:
<u>סעיף 1:</u>
נסתכל על התמונות שנוצרו ב־ VAE

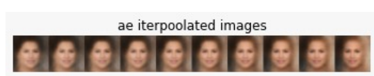


אפשר לראות שהתמונות שה־ VAE מייצר בקצוות זהות אבל התמונות שנוצרות בדרך לא דומות לאף ספרה ואין להם משמעות אמיתית.
נסתכל על התמונות שנוצרו ב־ GAN



אפשר לראות שה־ GAN כן משמר שכל תמונה בדרך היא גם נראת כמו ספרה.

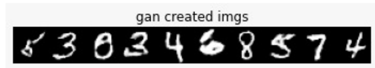
יתרה מזאת, אם נסתכל על עוד תמונות של VAE



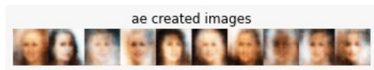
אפשר גם לשים לב שהתמונות מה־ VAE הן די מטושטשות כי המטרה שלו היא לשחזר את התמונה שתראה דומה למקורית בכל פיקסל ולא ערך סמנטי אחר.

צ"ל: ללמד VAE ולדגום
הוכחה:
<u>סעיף 2:</u>
נסתכל על תמונות שה־ VAE ייצר

אפשר לזהות ספרות בתמונות אך הן לא נראות הכי אמיתיות כי בשום שלב לא דרשנו שכל תמונה במרחב מדגם תראה אמיתית אלא רק אלא שעשינו להם $encoding$.
לעומתם אפשר לראות ש־ GAN מייצר



פה אפשר גם לזהות ספרות אך הם גם נראות ראליסטיות!
בנוסף לכך יש לנו תוצאות של VAE על $celebA$ אך הן יצאו לא כאלה טובות (אולי הוקטור לא היה מספיק או שלא רץ מספיק זמן)



אפשר שוב לראות את הטשטוש בתמונות של ה־ VAE אבל ניתן לראות שהוא כן מצליח לייצר פנים.

מ.ש.ל.ב.☺

2. **חלק תיאורטי:**

(א) **צ"ל:** כמה אילוצים מתמטיים כל תמונה נותנת?
הוכחה:
נסמן ב־ $\mathcal{X}$ את התמונות ב־ $ImageNet$ וב־ $\mathcal{Y}$ את הלייבלים ב־ $ImageNet$.
נתון שיש 1000 מחלקות ולכן $|\mathcal{Y}| = 1000$, נסמן בלי הגבלת הכלליות $\mathcal{Y} = \{1, \ldots, 1000\}$,
נסמן ב־ f את רשת הנוירונים.
נזכר שבבעיות קלסיפיקציה הרשת f מקבלת $x \in \mathcal{X}$ ומחזירה $y \in \mathbb{R}^{|\mathcal{Y}|}$ (וקטור הסתברות של הקלאסים).
לכן בהינתן דגימה $(x, y) \in \mathcal{X} \times \mathcal{Y}$, נרצה ש־ $f(x) = e_y$.
נשים לב כי נרצה ש־ $1 \leq i \leq |\mathcal{Y}|$ יתקיים $[f(x)]_i = [e_y]_i$,
ולכן יש $|\mathcal{Y}|$ אילוצים, במקרה שלנו מתקיים $|\mathcal{Y}| = 1000$ ולכן יש 1000 אילוצים.

מ.ש.ל.א.☺

(ב) **צ"ל:** נגזרות ביחס לקרנל וביחס לקלט באופן יעיל
הוכחה:
נסמן ב־ M_f את המטריצה המייצגת את פעולת הקונבולוציה עם וקטור f ולכן $C(I) = I * f = M_f \cdot I$.

$$\frac{\partial [C(I)]_k}{\partial [M_f]_{i,j}} = \frac{\partial [M_f \cdot I]_k}{\partial [M_f]_{i,j}} = \frac{\partial \left(\sum_m (M_f)_{k,m} \cdot I_m\right)}{\partial [M_f]_{i,j}} = \delta_{i=k} \frac{\partial \left(\sum_m (M_f)_{i,m} \cdot I_m\right)}{\partial [M_f]_{i,j}} = I_{j,i}$$

$$\frac{\partial \left((M_f)_{i,j} \cdot I_j\right)}{\partial [M_f]_{i,j}} = \delta_{i=k} \cdot I_j$$

לכן נקבל כי

$$\left[\frac{\partial L}{\partial C(I)} \cdot \Delta (M_f \cdot I)\right]_{i,j} = \left[\left[J_{M_f}(M_f \cdot I) \cdot \left[\frac{\partial L}{\partial C(I)}\right]^T\right]^T\right]_{i,j} = \left[J_{M_f}(M_f \cdot I) \cdot \left[\frac{\partial L}{\partial C(I)}\right]^T\right]_{j,i}$$
$$= \left[\sum_k \delta_{i=k} \cdot I_j \cdot \left[\frac{\partial L}{\partial C(I)}\right]_k\right]_{i,j} = \left[\frac{\partial L}{\partial C(I)}\right]_i \cdot I_j = I^T \cdot \frac{\partial L}{\partial C(I)}$$

כלומר לעדכון הגראדיאנטים לפי M_f צריך לחשב את $I^T \cdot \frac{\partial L}{\partial C(I)}$ שלוקח $O(n^2)$ זמן ריצה אבל אפשר לעשות ב־ $O(n)$ זיכרון על ידי חישוב איבר, עדכון ואז חישוב האיבר הבא.
עתה נחשב כי

$$\frac{\partial [C(I)]_i}{\partial [I]_j} = \frac{\partial [M_f \cdot I]_i}{\partial [I]_j} = \frac{\partial \left(\sum_{m=1}^{s} (M_f)_{i,m} \cdot I_m\right)}{\partial [I]_j} = [M_f]_{i,j}$$

ולכן נקבל כי $\Delta\left(M_f\cdot I\right)=\left[M_f\right]^T \implies J_I\left(M_f\cdot I\right)=\left[M_f\right]$.
עתה לחישוב הגראדיאנט לפי המשתנים I, צריך לחשב

$$\frac{\partial L}{\partial C\left(I\right)}\cdot\Delta\left(M_f\cdot I\right)=\frac{\partial L}{\partial C\left(I\right)}\cdot\left[M_f\right]^T=\left(M_f\cdot\left[\frac{\partial L}{\partial C\left(I\right)}\right]^T\right)^T$$

כלומר קיבלנו שבשביל לחשב את הגראדיאנטים של I מספיק לחשב את הקובנולוציה של f עם $\left[\frac{\partial L}{\partial C(I)}\right]^T$ ואז לעשות טרנספוז.
חישוב מכפלה של מטריצה בוקטור לוקחת $O\left(n^2\right)$ ועם האבחנה זה ייקח $O\left(n\log\left(n\right)\right)$ ולא דורש לשמור את כל M_f בזיכרון שמהווה ייעול משמעותי.

מ.ש.ל.ב.☺

(ג) **פתרון:**

i. **צ"ל:** מה זה FID, מה זה IS, מה הוא ה־ truncation trick?
הוכחה:
ה־ truncation trick הוא דגימה מההתפלגות שנבחרה אך זריקה של איברים שנדגמו בקצוות של הקטע שהוגדר.
הסיבה לכך היא שבזמן האימון, רואים קצוות אלה פחות ולכן התוצאות ייצאו פחות טובות ולכן הם בחרו לזרוק את הקצוות (לדגום מקטע קטן יותר).

ה־ IS הוא לוס שמריץ את ה־ $inceptionModule$ שמייצר התפלגות על קלאסים.
התמונות הנמדדות פה הן רק תמונות שה־ GAN ייצר.
בהתחלה נמדדת התפלגות שולית על הרבה דגימות.
לאחר מכן התפלגות של כל תמונה בפני עצמה ורוצים למקסם את השוני בין ההתפלגויות האלה.
הסיבה לכך היא שזה שמראה שיש הרבה שונות בתמונות שיוצרו ולא רק תמונה אחת ממש טובה. (המקסום נעשה על ידי KL Divergence שמודד מרחק בין התפלגויות).

ה־ FID הוא לוס שגם הוא מריץ את ה־ $inceptionModule$ שמייצר התפלגות על קלאסים.
הוא מודד פיצ'רים ב־ $inceptionModule$ על הרבה תמונות ומחשב תוחלת ושונות.
תחילה מחשבים את התוחלת והשונות לתמונות האמיתיות שנסמנן ב־ μ_{real},Σ_{real}.
לאחר מכן נעשה אותו התהליך לתמונות שה־ GAN ייצר ונסמנן ב־ μ_{fake},Σ_{fake}.
ה־ FID רוצה למזער את המרחק בין ההתפלגויות האלה על ידי הלוס

$$\left\|\mu_{real}-\mu_{fake}\right\|^2+\mathrm{tr}\left(\Sigma_{real}+\Sigma_{fake}-2\cdot\left(\Sigma_{real}\cdot\Sigma_{fake}\right)^{\frac{1}{2}}\right)$$

מ.ש.ל.ג.☺