

תיאור תהליך המחקר Improved SinGAN

רקע

מסמך זה נכתב ע"י הראל מנדלמן.

במסמך זה אנסה לתאר בקצרה את תהליך המחקר והתוצאות שקיבלתי כאשר ניסיתי למצוא שיפורים אפשריים לארכיטקטורת SinGAN במהלך חודש מאי 2023.

המחקר נעשה כפרויקט אישי למטרות לימודיות. על כן, לא בדקתי האם הניסויים/בדיקות שעשיתי כבר נעשו במחקרים אחרים באופן כללי ובהקשר של SinGAN בפרט. לא עשיתי סקר ספרות ולא בדקתי אילו שיפורים נעשו ל SinGAN עד היום במחקרי המשך. המחקר נעשה תוך מיקוד בעיקר במשימה של Super Resolution, אם זאת יתכן כי התוצאות שקיבלתי רלוונטיות גם לאפליקציות אחרות.

המסמך נכתב בעברית משום שהוא מסמך שנכתב תוך כדי תהליך המחקר ולא אמור לשמש לצרכים חיצוניים בשלב זה.

קישור ל repository עם הקוד שבו מימשתי את הניסויים השונים שעשיתי:

https://github.com/harel147/improved_singan

לטובת הצגה מסודרת של התוצאות שקיבלתי בניתי אתר פשוט שמסכם את התוצאות של כל ניסוי. בכל הניסויים השתמשתי ב 3 תמונות קבועות, התמונה "stone face" שמשתמשים בה ב repository המקורי עבור הצגה של SR, והתמונות "cows" ו-"colosseum" שמשתמשים בהן ה repository המקורי עבור אימון רגיל, מהן חתכתי תמונה בגודל של בערך חצי/רבע מהתמונה המקורית. החיתוך התבצע במטרה להבליט את תוצאות ה-SR. קישור לאתר התוצאות:

https://harel147.github.io/html_research_results/index.html

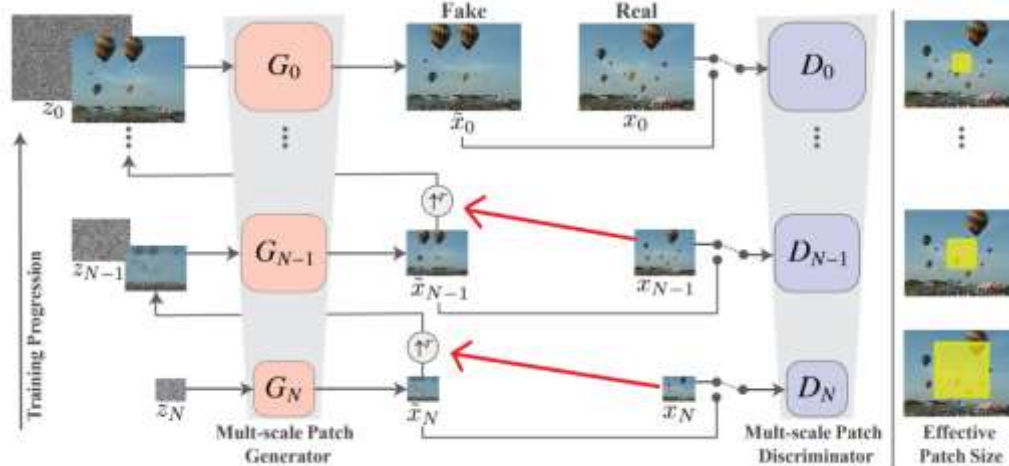
תוכן עניינים

1. כיווני מחשבה שעלו לי לאחר קריאת המאמר ומעבר ראשוני על המימוש.
2. בחינת דרכים לקיצור תהליך האימון.
3. שיפור Super Resolution באמצעות שימוש בתמונה המקורית בתהליך האימון.
4. מציאת גישה טובה יותר לעשות upsample בתהליך האימון.

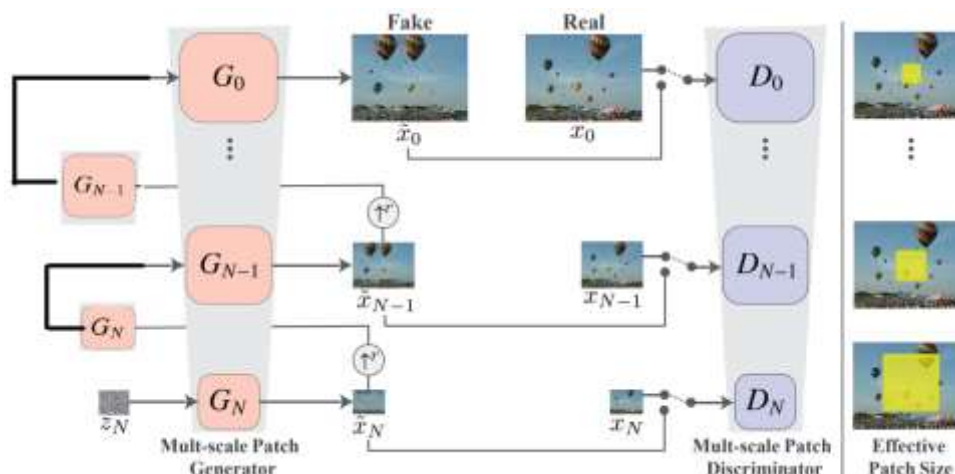
כיווני מחשבה שעלו לי לאחר קריאת המאמר ומעבר ראשוני על המימוש

אלה הם הכיוונים שבחנתי בתהליך המחקר, פירוט של כל אחד מהם, הצעדים שביצעתי והתוצאות שקיבלתי בהמשך המסמך:

1. האימון של כל scale הוא באופן דיפולטי 2000 epoch'ים, ולוקח הרבה זמן לאמן מודל לתמונה בגלל זה. לבחון האם אפשר למצוא דרך טובה יותר לבחור את מספר ה epoch'ים לכל scale.
2. אינטואיטיבית נשמע לי הגיוני שספציפית עבור המשימה של SR, שימוש בתמונה המקורית בתהליך האימון זה דבר שיכול לעזור. במילים אחרות, לעשות קצת "overfit" על התמונה האמיתית. כלומר, כרגע נראה שהתמונות שנכנסות לכל generator הן תמיד תמונות שג'ונרטו דרך כל ה scale'ים שנמצאים מתחתיו, ונראה לי ניסוי ששווה לעשות זה להכניס את התמונה עצמה (במינון כלשהו, אולי רק ב scale האחרון?) כ Input ל generator. כלומר באופן הבא (חצים אדומים):



3. מציאת גישה טובה יותר לעשות את תהליך ה upsample. זה נראה שלאופן שבו נעשה ה upsample תהיה השפעה רבה על תוצאות האימון.
 - a. להשתמש ברשת שתלמד דרך טובה יותר לעשות upsample בכל scale בהתאם לתמונה הנלמדת?
 - b. לעשות "upsample" באמצעות SR? אפשר כנראה לקבל את זה בחינם כי גם ככה אנחנו מאמנים את כל ה scale'ים. כלומר:



4. לימוד של "אופי" האיזורים השונים בתמונה, ואז כאשר עושים SR שימוש לא רק ב generator של ה scale האחרון, אלה שימוש גם בג'נרטורים ב scale'ים נמוכים יותר בהתאם "לאופי" של כל איזור בתמונה.

קיצור תהליך האימון

1. לטובת קבלת אינטואיציה, מימנתי מספר כלי debug מרכזיים:

- a. שמירה גם של תמונות ביניים במהלך תהליך האימון של כל scale ולא רק של התוצאה הסופית, כלומר שמירת fake ו- $G(z)$ כל 200 epoch'ים. המטרה היא להבין באימון של כל scale מתי הוא הגיע "לרוויה" ולנסות להבין האם יש חוקיות כלשהי או תופעות שחוזרות על עצמן בתמונות שונות.
- b. שמירה של קובץ ששומר את זמני האימון של כל scale והזמן הכולל של האימון. לדוגמה:

```
1 level 0 time: 1.91 minuets
2 level 1 time: 2.29 minuets
3 level 2 time: 2.27 minuets
4 level 3 time: 2.18 minuets
5 level 4 time: 2.57 minuets
6 level 5 time: 2.66 minuets
7 level 6 time: 3.61 minuets
8 level 7 time: 6.18 minuets
9 total time: 23.67 minuets
```

c. הדפסת שגיאת השחזור לאחר כל 200 epoch'ים במהלך תהליך האימון.

2. הרעיון הראשון שעלה לי כאשר קראתי את המאמר, עבור קיצור זמן האימון, היה לעשות סוג של "transfer learning" בין ה scale'ים השונים, כלומר שכל scale לא יאותחל עם משקולות רנדומליות אלה יתחיל מה- check point האחרון של ה scale הקודם. כאשר עברתי על הקוד גיליתי שהפרקטיקה הזאת ממומשת בקוד למרות שלא מוזכרת במאמר ולכן זנחתי את הכיוון הזה.

3. לאחר מכן ניסיתי באמצעות שימוש בכלי ה- debug שהזכרתי קודם להבין האם קיימת חוקיות מתי האימון של כל scale "מתנוון". תופעות שזיהיתי:

- a. כאשר בחנתי את התוצאות שמקבלים ל $G(z)$ אחרי כל 200 epoch'ים:
- i. ב- scale הראשון – מקבלים תמונה לא ברורה לאורך כל האימון, גם ב 200 epoch וגם ב 2000, הייתי אומר שהתמונה שקולה בכמה שהיא לא ברורה בשני המקרים.
- ii. ב scale השני – נראה שה- epoch האחרון (2000) הכי טוב (יותר טוב מהתמונה המתקבלת עבור epoch'ים 1600, 1800, ... 200).
- iii. החל מה- scale השלישי – אני לא מצליח לראות בעין הבדל בין התמונה שנוצרה ב epoch ה 200 לבין התמונה שנוצרה ב 2000 epoch.
- b. כאשר בחנתי את התוצאות שמקבלים ל fake אחרי כל 200 epoch'ים:
- i. באופן כללי נראה שהפער יותר נוכח בין אימון עם 200 epoch'ים לאימון מלא מאשר עבור התוצאות של $G(z)$. עם זאת, נראה שזה מאוד תלוי בתמונה ויש תמונות שעבורן כן מתקבלת תמונה טובה גם עם מעט epoch'ים.

4. בהמשך להבחנות האלה, החלטתי לבדוק האם יתכן שקבלת $G(z)$ טוב הוא תנאי מספיק עבור עצירת האימון של ה scale ומעבר לאימון ה scale הבא. על כן, מימנתי אפשרות לבצע אימון עם תנאי עצירה עבור כל ה scale'ים חוץ מאשר ה scale האחרון כאשר התנאי הוא שאם ה- epoch גדול מ 200, ושגיאת השחזור קטנה מ 0.025 יש לעצור את אימון ה scale ולעבור לאימון ה scale הבא. אימון באופן הזה גרם לכך שאם לדוגמה מאמנים 8 scale'ים

עבור תמונה מסויימת, אז אימון ה scale הראשונה לא יעצר לפני 2000 epoch'ים או יעצר קרוב למספר הזה (שכן ב scale הראשון שגיאת השחזור לא מגיעה ל 0.025 ב 2000 epoch'ים), אימון ה scale השני ייעצר לאחר בערך 1500 epoch'ים, scale'ים 3 עד 7 יעצרו אחרי 200-400 epoch'ים בלבד, ואת ה scale האחרון אני בכוונה מאמן אימון מלא של 2000 epoch'ים. תנאי העצירה מקצר את זמני האימון באופן משמעותי משום שכלל שה scale יותר גדול גם זמן האימון גדול יותר, וב scale'ים הגדולים האימון ירד לבערך 200 epoch'ים כי מגיעים מהר לשגיאת שחזור קטנה.

5. ניתן לראות את התוצאות שקיבלתי עם שימוש בתנאי העצירה על שגיאת השחזור עבור SR מול תוצאות האימון המקורי באתר התוצאות [בעמוד הזה](#). ניתן לראות לפי מדידת זמני האימון של כל scale שזמני האימון פחתו באופן משמעותי, לדוגמה עבור אימון התמונה החתוכה "cows" עבור SR:

1 level 0 time: 2.27 minuets	1 level 0 time: 2.1 minuets
2 level 1 time: 2.38 minuets	2 level 1 time: 1.66 minuets
3 level 2 time: 2.98 minuets	3 level 2 time: 0.69 minuets
4 level 3 time: 2.93 minuets	4 level 3 time: 0.49 minuets
5 level 4 time: 3.22 minuets	5 level 4 time: 0.54 minuets
6 level 5 time: 3.84 minuets	6 level 5 time: 0.39 minuets
7 level 6 time: 4.92 minuets	7 level 6 time: 0.53 minuets
8 level 7 time: 7.76 minuets	8 level 7 time: 7.38 minuets
9 total time: 30.3 minuets	9 total time: 13.78 minuets

Original training

Early stop training

זמן האימון של ה scale הראשון דומה, אבל בשאר ה scale'ים זמן האימון יורד משמעותית, חוץ מאשר ב scale האחרון שבו אילצתי כמות epoch'ים מלאה.

6. לא הגעתי למסקנה חותכת בשלב זה בנוגע לאפקטיביות של תנאי העצירה המוצע. ניתן לראות שניתן להגיע לתוצאות סבירות בשימוש בתנאי העצירה הנ"ל, לעיתים התמונות אף נראות יותר טבעיות כאשר משתמשים בו. יתכן כי ניתן למצוא תנאי עצירה מורכב יותר, שמשתמש גם בשגיאת השחזור על מנת לקצר את תהליך האימון בצורה אפקטיבית יותר.
7. בשלב זה בחרתי לזנוח את הכיוון הזה ולבחון כיוונים נוספים.

שיפור SR באמצעות ניצול התמונה המקורית בתהליך האימון

1. כמו שצינתי קודם ברשימת כיווני המחקר בתחילת המסמך, האינטואיציה שלי הייתה שעבור המשימה של SR יש היגיון בלעשות "overfit" של האימון על התמונה האמיתית, ולא לאמן רק על בסיס תמונות שמג'נרטות דרך כל הפירמידה.
 2. על מנת לאפשר גמישות בשימוש בתמונה המקורית במהלך תהליך האימון, שתקל עלי בעת ביצוע הניסויים השונים שתכננתי לעשות, הכנסתי בקוד את השינויים הבאים:
 - a. בפונקציה draw_concat היו קיימים שני mode'ים: 'rand' ו-'rec'. הפונקציה הזו אחראית על הכנת התמונות עבור האימון של ה scale הנוכחי וביצוע upsample על ידי ג'נרט התמונות דרך כל ה scale'ים הקודמים שכבר אומנו. על מנת לאפשר אימון על בסיס התמונה האמיתית, הוספתי לפונקציה שני mode'ים חדשים שניתן להפעיל: 'real_train' ו-'real_train_rec'. ה mode'ים האלה מאפשרים לקחת את התמונה האמיתית מהשכבה הקודמת להוסיף לה רעש ולעשות לה upsample, במקום לג'נרט תמונה דרך כל ה scale'ים הקודמים. 'real_train' עבור השגיאה הרגילה ו-'real_train_rec' עבור שגיאת השחזור.
 - b. הכנסה של תנאי עצירה שמאפשר לעצור מיידית (אחרי 2 epoch'ים בלבד) את האימון של כל ה scale'ים התחתונים ובעצם לאמן רק את ה scale האחרון. הסיבה שהיה צורך בתנאי זה היא שבחלק מהניסויים שאני אתאר בהמשך אין ערך באימון ה scale'ים התחתונים (למעט ה "transfer" של המשקולות בין scale ל scale).
- פירוט על כך בהמשך.

- c. אפשרות לבחור את כל אחד מה-mode'ים, המקוריים ואלה שאני הוספתי, בנפרד עבור כל ה-scale'ים התחתונים, ועבור ה scale האחרון. כלומר, ניתן עכשיו לדוגמה לקבוע שה-scale'ים התחתונים יאומנו ב mode כמו 'rand' וה-scale העליון יאומן על התמונה המקורית עם 'real_train', כנ"ל עבור שגיאת השחזור.
3. אלה הניסויים שביצעתי ובהם השתמשתי גם בתמונה המקורית בתהליך האימון, מתחת לטבלה ניתן למצוא פירוט של כל ניסוי. כל התוצאות שקיבלתי מופיעות באתר התוצאות תחת הכותרת "train with real image experiments". ניתן להגיע לתוצאות של כל ניסוי ע"י לחיצה עליו.

ניסוי	אימון מלא של כל ה scale'ים או אימון של ה scale'ים התחתונים עם 2 epoch'ים (והאחרון מלא)	אימון השגיאה הרגילה ב-scale'ים התחתונים	אימון השגיאה הרגילה ב-scale האחרון	אימון שגיאת השחזור ב-scale'ים התחתונים	אימון שגיאת השחזור ב-scale האחרון
a	2 epoch'ים	N/A	real_train	N/A	rec
b	מלא	real_train	real_train	rec	rec
c	מלא	rand	real_train	rec	rec
d	2 epoch'ים	N/A	real_train	N/A	real_train_rec
e	מלא	real_train	real_train	real_train_rec	real_train_rec
f	מלא	rand	real_train	rec	real_train_rec
g	מלא	rand	rand	real_train_rec	real_train_rec

** N/A כאשר ה scale'ים התחתונים לא אומנו ולכן ה-mode לא רלוונטי.

4. פירוט הניסויים:

- a. אימון רק של ה scale האחרון (כל ה scale'ים התחתונים רק 2 epoch'ים) עם התמונה האמיתית לשגיאה הרגילה ('real_train') ועם התמונה המג'ונטרת דרך כל ה scale'ים ('rec') לשגיאת השחזור. כמובן שלא סביר לקבל תוצאות טובות במקרה הזה משום שמג'נרטים תמונות דרך generator'ים שלא אומנו, ואכן מקבלים תוצאות לא טובות.
- b. אימון מלא של כל ה scale'ים, שימוש בתמונה האמיתית לשגיאה הרגילה ('real_train') ובתמונה המג'ונטרת דרך כל ה scale'ים ('rec') לשגיאת השחזור. כלומר, עבור השגיאה הרגילה מאמנים עם התמונה המקורית, ועבור שגיאת השחזור מאמנים באותו אופן כמו באימון המקורי. נראה שמקבלים תוצאות לא רעות.
- c. אימון מלא של כל ה scale'ים, שימוש בתמונה המג'ונטרת בכל ה scale'ים ('rand') לשגיאה הרגילה בדומה לאימון המקורי חוץ מאשר ב scale האחרון שבו מאמנים עם התמונה המקורית האמיתית לשגיאה הרגילה ('real_train'). בכל ה scale'ים כולל האחרון אימון עם התמונה המג'ונטרת דרך כל ה scale'ים ('rec') לשגיאת השחזור. כלומר, אימון רגיל רק שאת ה scale האחרון מאמנים עם התמונה המקורית עבור השגיאה הרגילה. תוצאות טובות.
- d. אימון רק של ה scale האחרון (כל ה scale'ים התחתונים רק 2 epoch'ים) עם התמונה האמיתית לשגיאה הרגילה ('real_train') ועם התמונה האמיתית גם לשגיאת השחזור ('real_train_rec'). הסיבה לא לאמן את ה scale'ים התחתונים כי עקרונית לא נעשה בהם שימוש באימון של ה scale האחרון. הערך היחיד של כן לאמן כמות epoch'ים מלאה היא בשביל "transfer" למשקלים בין ה scale'ים. כלומר, באימון הזה אני מאמן רק את ה scale האחרון ועושה את זה רק על בסיס

התמונה האמיתית ולא על בסיס תמונה שמג'ונרטת דרך כל ה scale'ים. מקבלים תוצאות קצת מוזרות אבל מעניינות.

e. שוב אימון רק על בסיס התמונה האמיתית גם עבור השגיאה הרגילה ('real_train') וגם עבור שגיאת השחזור ('real_train_rec'), אבל הפעם כן לעשות אימון מלא לכל ה scale'ים (2000 epoch'ים) בשביל אולי להרוויח משהו מהעברת המשקלים מ scale ל scale. מעניין לציין שנראה שמתקבלות תוצאות פחות טובות מאשר האימון שבו לא אימנתי את ה scale'ים התחתונים (ניסוי d).

f. אימון רגיל בכל ה scale'ים התחתונים, שימוש בתמונה המג'ונרטת לשגיאה הרגילה ('rand') ובתמונה המג'ונרטת לשגיאת השחזור ('rec'). ורק ב scale האחרון שימוש בתמונה האמיתית לשגיאה הרגילה ('real_train') ולשגיאת השחזור ('real_train_rec'). כלומר, ה scale'ים התחתונים מאומנים כרגיל וה scale האחרון מאומן על התמונה האמיתית. גם במקרה הזה בגלל שבפועל אני מאמן את ה scale האחרון רק על התמונה האמיתית אז לכאורה אין סיבה לאמן את ה scale'ים התחתונים, אבל אני עושה את זה בכל זאת בשביל רוח אפשרי מה- "transfer" של המשקולות מה- scale'ים התחתונים. מקבלים תוצאות קצת מוזרות אבל מעניינות.

g. אימון של כל ה scale'ים כולל האחרון עם התמונה המג'ונרטת עבור השגיאה הרגילה ('rand'), אבל אימון של כל ה scale'ים כולל האחרון עבור שגיאת השחזור עם התמונה האמיתית ('real_train_rec'). תוצאות טובות, אולי אפילו מאוד טובות.

h. ניתן לבחון השפעות של שילובים נוספים שלא הגעתי ללנסות, כמו אימון שבכל scale'ים משתמש גם בתמונה האמיתית וגם במג'ונרטת, לדוגמה בכל epoch לאמן חלק מהאיטרציות על התמונה האמיתית וחלק מהאיטרציות דרך התמונה המג'ונרטת.

5. גם פה לא הגעתי לתוצאות חד משמעיות. אני חושב שהתוצאות מאוד מעניינות, ועל ידי בדיקה על דאטאסט גדול יותר יתכן כי ניתן יהיה למצוא קומבינציה אשר נותנת תוצאות טובות יותר מאשר האימון המקורי. בנוסף, יתכן כי ניתן לאפיין עבור אילו תמונות עדיף להשתמש באיזו קומבינציה.
6. זו לדוגמה התוצאה של האימון המקורי מצד שמאל מול התוצאה של ניסוי g מצד ימין:



מצאת גישה טובה יותר לעשות upsample

- להשתמש ברשת שתלמד דרך טובה יותר לעשות upsample מאשר הגישה הנאיבית?
a. לא יצא לי לבחון את הרעיון הזה יותר לעומק.
- שימוש ב SR במהלך האימון עצמו במקום לעשות upsample.
a. תיאור הניסויים בפרק הבא.

שימוש ב SR במהלך האימון עצמו במקום לעשות upsample

- האינטואיציה מאחורי הרעיון הזה הייתה שיש במהלך האימון יכולת מובנת "להגדיל" את התמונה באמצעות SR, אז למה לא להכין את התמונה עבור ה scale הבא באמצעות SR במקום באמצעות upsample נאיבי?
- גם במקרה הזה, אני יכול להכין תמונה עבור ה scale הבא גם עבור השגיאה הרגילה וגם עבור שגיאת השחזור, ועכשיו שכבר מימשתי mode של אימון על התמונה האמיתית ('real_train' ו 'real_train_rec'), אז ניתן לשלב גם בהם אפשרות לאמן עם SR בתור טכניקת upsample. לכן, בחרתי להוסיף במימוש יכולת להעביר את הפרמטרים "upsample_type" ו- "rec_upsample_type" ששולטים על טכניקת ה upsample עבור השגיאה הרגילה ועבור שגיאת השחזור, ללא תלות באיזה mode משתמשים 'rand', 'real_train', 'rec' או 'real_train_rec'. גם כאן הוספתי שליטה נפרדת על ה scale של התחתונים וה- scale האחרון, כלומר ניתן לדוגמה להשתמש ב SR כטכניקת upsample רק ב scale האחרון.
- אלה הניסויים שביצעתי ובהם השתמשתי ב- SR כטכניקת upsample, הניסויים משלבים גם קומבינציות שונות של האימון על התמונה המקורית כפי שתיארתי קודם. כל התוצאות שקיבלתי מופיעות באתר התוצאות תחת הכותרת "train with SR for upsample experiments". ניתן להגיע לתוצאות של כל ניסוי ע"י לחיצה עליו.

ניסוי	טכניקת upsample ב-scale התחתונים עבור השגיאה הרגילה	טכניקת upsample ב-scale האחרון עבור השגיאה הרגילה	טכניקת upsample ב-scale התחתונים עבור שגיאת השחזור	טכניקת upsample ב-scale האחרון עבור שגיאת השחזור	אימון השגיאה הרגילה ב-scale האחרון	אימון שגיאת השחזור ב-scale התחתונים	אימון שגיאת השחזור ב-scale האחרון
a	sr	sr	regular	regular	rand	real_train_rec	real_train_rec
b	sr	sr	sr	sr	rand	rec	rec
c	regular	sr	regular	regular	rand	rec	rec
d	regular	sr	regular	sr	rand	rec	real_train_rec
e	regular	sr	regular	sr	rand	rec	real_train
f	sr	sr	regular	regular	real_train	real_train_rec	real_train_rec
g	regular	sr	regular	sr	real_train	real_train_rec	real_train_rec
h	regular	sr	regular	regular	real_train	real_train_rec	real_train_rec
i	regular	sr	regular	regular	real_train	rec	rec

- באופן כללי, לא הצלחתי לקבל תוצאות טובות באמצעות שימוש ב SR בתור טכניקת upsample. הניסוי היחיד שנתן תוצאות סבירות יחסית הוא ניסוי i, ולכן ביצעתי אותו על שלושת התמונות ולא רק על "stone face".
- נראה שהשימוש ב sr כטכניקת upsample ב scale של התחתונים הורס את האימון כמו שניתן לראות בניסויים a, b, f. יתכן שהבעיה נובעת מכך שעושים את ה upsample ע"י SR כבר מה- scale הראשון ואם זה יתבצע החל מה- scale השני או השלישי אז לא יתקבל

אפקט דומה. עם זאת, גם כאשר נעשה upsample ע"י SR רק ב scale האחרון עדין נראתה פגיעה בתוצאות.

6. יתכן שהסיבה לפגיעה בתוצאות בשימוש ב SR עבור ה upsample היא שביצוע SR באיזשהו אופן בעצם "מפחית" את הרעש שנוסף לתמונה, ויתכן שהרעש דווקא עוזר לאימון, ולכן דווקא השימוש ב SR והעברה של תמונה "פחות מורעשת" ל scale הבא דווקא פוגע באימון בניגוד למה שחשבתי. יתכן שפתרון לכך יכול להיות הוספה של רעש לאחר ביצוע ה SR, כלומר ביצוע upsample והוספת רעש -> ביצוע SR -> הוספה של רעש עוד פעם -> העברה ל scale הבא. לא בחנתי זאת בשלב זה.