

## תיאור תהליך המחקר Improved SinGAN

### רקע

מסמך זה נכתב ע"י הראל מנדלמן.

במסמך זה אנסה לתאר את תהליך המחקר והתוצאות שקיבלתי כאשר ניסיתי למצוא שיפורים אפשריים לארכיטקטורת SinGAN.

המחקר נעשה כפרויקט אישי למטרות לימודיות, ועל כן התמקד ספציפית בארכיטקטורת SinGAN ובתוצאות שלה, ולא התבצעה השוואה מול מחקרים אחרים. המחקר נעשה תוך מיקוד במשימה של Super Resolution, עם זאת יתכן כי התוצאות שהתקבלו 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 של הפרויקט:

[https://github.com/harel147/improved\\_singan](https://github.com/harel147/improved_singan)

לטובת הצגה מסודרת של התוצאות בניתי אתר פשוט שמסכם את התוצאות של כל ניסוי. בכל הניסויים השתמשתי ב 3 תמונות קבועות, התמונה "stone face" שמשתמשים בה במאמר המקורי המקורי עבור הצגה של SR, והתמונות "cows" ו-"colusseum" שמשתמשים בהן במאמר 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:

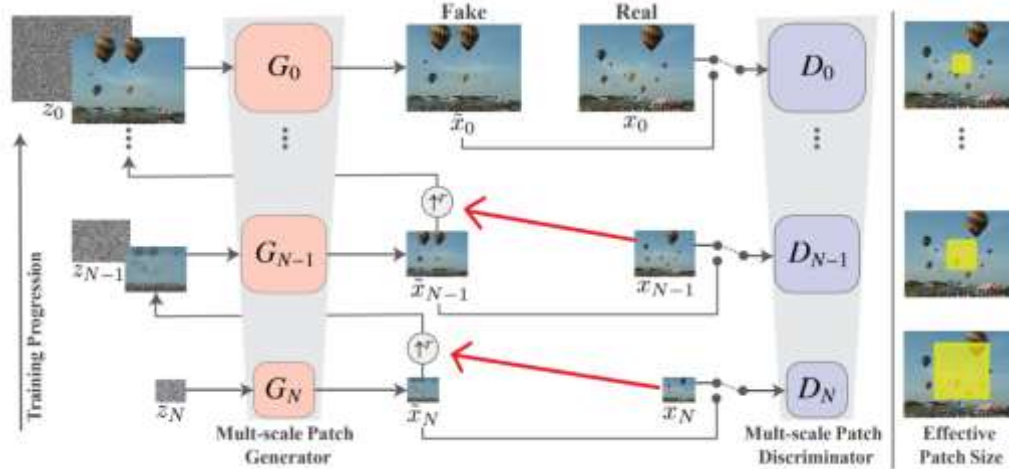
[https://harel147.github.io/html\\_research\\_results/index.html](https://harel147.github.io/html_research_results/index.html)

### תוכן עניינים

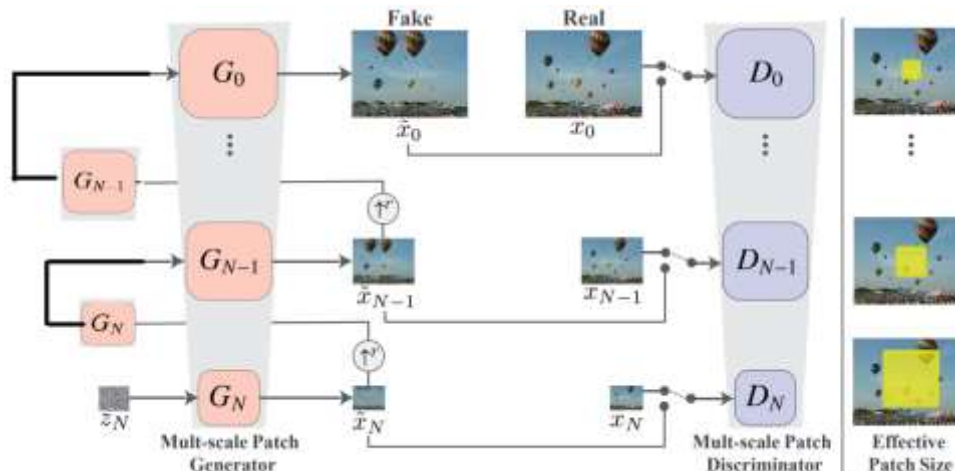
1. כיווני מחשבה לאחר קריאת המאמר ומעבר ראשוני על המימוש הקיים.
2. בחינת דרכים לקיצור תהליך האימון.
3. שיפור Super Resolution באמצעות שימוש בתמונה המקורית בתהליך האימון.
4. מציאת גישה טובה יותר לעשות Upsample בתהליך האימון.

## כיווני מחשבה לאחר קריאת המאמר ומעבר ראשוני על המימוש

1. האימון של כל scale מוגדר באופן דיפולטי ל- 2000 epoch'ים, ועל כן לוקח זמן רב לאמן מודל לתמונה. כדאי לבחון האם אפשר למצוא דרך טובה יותר לבחור את מספר ה epoch'ים לכל scale.
2. אינטואיטיבית, עבור המשימה של SR, שימוש בתמונה המקורית בתהליך האימון יכול לעזור. במילים אחרות, לעשות קצת "overfit" על התמונה האמיתית. כרגע, התמונות שנכנסות לכל generator הן תמיד תמונות שג'ונרטו דרך כל ה scale'ים שנמצאים מתחתיו. אני מאמין שלהכניס את התמונה עצמה כקלט ל generator, במינון כלשהו, יכול להיות ניסוי מעניין. כלומר באופן הבא:



3. מציאת גישה טובה יותר לעשות את תהליך ה upsample. יתכן 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? כלומר באופן הבא:



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. עבור 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 שחשבתי לבחון 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 – מתקבלת תמונה לא ברורה לאורך כל האימון, גם ב 200 epoch וגם ב 2000, נראה שהתמונה שקולה בכמה שהיא לא ברורה בשני המקרים.
      - ii. ב scale השני – נראה שה- epoch האחרון (2000) הוא הטוב ביותר בהשוואה לתמונות המתקבלות עבור epoch'ים 1600, 1800, ... 200.
      - iii. החל מה- scale השלישי – לא ניתן להבחין בהבדל בין התמונה שנוצרה ב epoch ה 200 לבין התמונה שנוצרה ב 2000 epoch.
    - b. כאשר בחנתי את התוצאות המתקבלות 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, למעט ב scale האחרון. 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 מסוימת:
- 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.
  - אימון scale'ים 3 עד 7 יעצרו אחרי 200-400 epoch'ים בלבד.
  - ה scale האחרון עובר במכוון אימון מלא של 2000 epoch'ים.
5. תנאי העצירה מקצר את זמני האימון באופן משמעותי משום שכל scale יותר גדול גם זמן האימון גדול יותר, וב- scale'ים הגדולים האימון ירד לבערך 200 epoch'ים כי מגיעים מהר לשגיאת שחזור קטנה.
6. ניתן לראות את התוצאות שהתקבלו 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 שמאומן באופן מכוון בכמות epoch'ים מלאה.

7. 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.

### שיפור SR באמצעות שימוש בתמונה המקורית בתהליך האימון

- כפי שצוין קודם ברשימת כיווני המחקר, האינטואיציה מאחורי ניסוי זה היא שעבור המשימה של SR יש היגיון בלעשות באימון "overfit" על התמונה האמיתית, ולא לאמן רק על בסיס תמונות שמג'נרטות דרך כל הפירמידה.
- על מנת לאפשר גמישות בשימוש בתמונה המקורית במהלך תהליך האימון, אשר תאפשר ביצוע של ניסויים שונים, מומשו השינויים הבאים בקוד המקורי של המאמר:
  - בפונקציה draw\_concat היו קיימים שני mode'ים: 'rand' ו-'rec'. פונקציה זו 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, התווספו 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.
  - הוספה 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 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, המקוריים ואלה שהתווספו, בנפרד עבור כל ה scale'ים התחתונים, ועבור ה scale האחרון. כלומר, ניתן קעת 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 שהתקבלו מופיעות באתר התוצאות בעמוד [Train with real image experiments](#). ניתן להגיע לתוצאות של כל ניסוי ע"י לחיצה על המזהה שלו בטבלה.

ניסוי	אימון מלא של כל ה scale'ים או אימון של ה scale'ים התחתונים עם 2 epoch'ים (ומאחרון מלא)	אימון הרגילה ב-scale'ים התחתונים	אימון הרגילה ב-scale האחרון	אימון שגיאת השחזור ב-scale'ים התחתונים	אימון שגיאת השחזור ב-scale האחרון
<a href="#">a</a>	2 epoch'ים	N/A	real_train	N/A	rec
<a href="#">b</a>	מלא	real_train	real_train	rec	rec
<a href="#">c</a>	מלא	rand	real_train	rec	rec
<a href="#">d</a>	2 epoch'ים	N/A	real_train	N/A	real_train_rec
<a href="#">e</a>	מלא	real_train	real_train	real_train_rec	real_train_rec
<a href="#">f</a>	מלא	rand	real_train	rec	real_train_rec
<a href="#">g</a>	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:

- 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 מתקבלות תוצאות לא טובות.
- 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.
- 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.
- 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. מתקבלות תוצאות קצת מוזרות, אך אם זאת לא רעות ביחס לזה שהן התקבלו באמצעות אימון של ה- scale האחרון בלבד.

e. 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 אומנו 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 אומנו כרגיל וה scale האחרון אומן על התמונה האמיתית. גם במקרה הזה בגלל שבפועל ה scale האחרון אומן רק על התמונה האמיתית אז לכאורה אין סיבה לאמן את ה scale'ים התחתונים, וזה נעשה בכל זאת בשביל רווח אפשרי מה- "transfer" של המשקולים מה- scale'ים התחתונים. מתקבלות 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 מאוד טובות.

5. ניתן לבחון השפעות של שילובים נוספים בתהליך האימון בין התמונה האמיתית לתמונה המג'ונרטת. לדוגמה, אימון שבכל scale משתמש גם בתמונה האמיתית וגם במג'ונרטת, כלומר בכל epoch לאמן חלק מהאיטרציות על התמונה האמיתית וחלק מהאיטרציות על התמונה המג'ונרטת.

6. גם כאן 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, או עבור איזה פרמטר של התמונה שרוצים למקסם עליו, עדיף להשתמש באיזו קומבינציה.

7. זו לדוגמה התוצאה של האימון המקורי מצד שמאל מול התוצאה של ניסוי g מצד ימין, שבו כל ה- scale אומנו עם התמונה המג'ונרטת עבור השגיאה הרגילה ועם התמונה האמיתית עבור שגיאת השחזור:





8. ואלה התוצאות של ניסוי d, שמעניין לראות בו שאמצעות אימון רק של ה scale האחרון עם התמונה האמיתית עבור שתי השגיאות, ללא אימון כלל של כל ה scale'ים התחתונים, ניתן להגיע לתוצאות לא רעות:



## שימוש ב SR בתור טכניקת upsample במהלך האימון

1. האינטואיציה מאחורי הרעיון הזה הייתה שיש במהלך האימון יכולת מובנת "להגדיל" את התמונה באמצעות SR. כלומר, להכין את התמונה עבור ה scale הבא באמצעות SR במקום באמצעות upsample נאיבי.
2. בדומה לאימון עם התמונה האמיתית, גם כאן ניתן להכין תמונה עבור ה scale הבא באמצעות SR כטכניקת upsample גם עבור השגיאה הרגילה וגם עבור שגיאת השחזור. יתר על כן, לאחר מימוש ה-mode של אימון על התמונה האמיתית ('real\_train' ו 'real\_train\_rec'), ניתן לשלב גם בהם אפשרות לאמן עם SR בתור טכניקת upsample. לכן, התווספה 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 התווספה שליטה נפרדת על ה scale'ים התחתונים וה- scale האחרון, כלומר ניתן לדוגמה להשתמש ב SR כטכניקת upsample רק ב scale האחרון.
3. אלה הניסויים שערכתי ובהם השתמשתי ב- SR כטכניקת upsample, הניסויים משלבים גם קומבינציות שונות של האימון על התמונה המקורית כפי שתיארתי קודם. כל התוצאות שהתקבלו מופיעות באתר התוצאות בעמוד [train with SR for upsample experiments](#). ניתן להגיע לתוצאות של כל ניסוי ע"י לחיצה עליו.

ניסוי	טכניקת upsample ב-scale'ים התחתונים עבור השגיאה הרגילה	טכניקת upsample ב-scale'ים התחתונים עבור השגיאה הרגילה	טכניקת upsample ב-scale'ים התחתונים עבור שגיאת השחזור	אימון השגיאה הרגילה ב-scale'ים התחתונים	אימון שגיאת השחזור ב-scale'ים התחתונים	אימון שגיאת השחזור ב-scale'ים התחתונים	אימון שגיאת השחזור ב-scale'ים התחתונים
<a href="#">a</a>	sr	sr	regular	rand	rand	real_train_rec	real_train_rec
<a href="#">b</a>	sr	sr	sr	rand	rand	rec	rec
<a href="#">c</a>	regular	sr	regular	rand	rand	rec	rec
<a href="#">d</a>	regular	sr	regular	rand	real_train	rec	real_train_rec
<a href="#">e</a>	regular	sr	regular	rand	real_train	rec	rec
<a href="#">f</a>	sr	sr	regular	real_train	real_train	real_train_rec	real_train_rec
<a href="#">g</a>	regular	sr	regular	real_train	real_train	real_train_rec	real_train_rec
<a href="#">h</a>	regular	sr	regular	real_train	real_train	real_train_rec	real_train_rec
<a href="#">i</a>	regular	sr	regular	real_train	real_train	rec	rec

4. באופן כללי, לא הצלחתי לקבל תוצאות טובות באמצעות שימוש ב SR בתור טכניקת upsample. הניסוי היחיד שנתן תוצאות סבירות הוא ניסוי i, ולכן ביצעתי אותו על שלושת התמונות ולא רק על "stone face".
5. נראה שהשימוש ב sr כטכניקת upsample ב scale'ים התחתונים פוגע באימון 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 בתור טכניקת 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 לתזה הראשונית. 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.