FIG SUPPLEMENTAL: Compression fracture: T1w, T2w and conventional acquired STIR (left) and DL synthesized STIR (far right). Note the improved IQ and reduced artifact of the synthetic image derived from the high SNR acquired T1w and T2w images.

FIG SUPPLEMENTAL: Compression fracture: T1w, T2w and conventional acquired STIR (left) and DL synthesized STIR (far right). Note the improved IQ and reduced artifact of the synthetic image derived from the high SNR acquired T1w and T2w images.

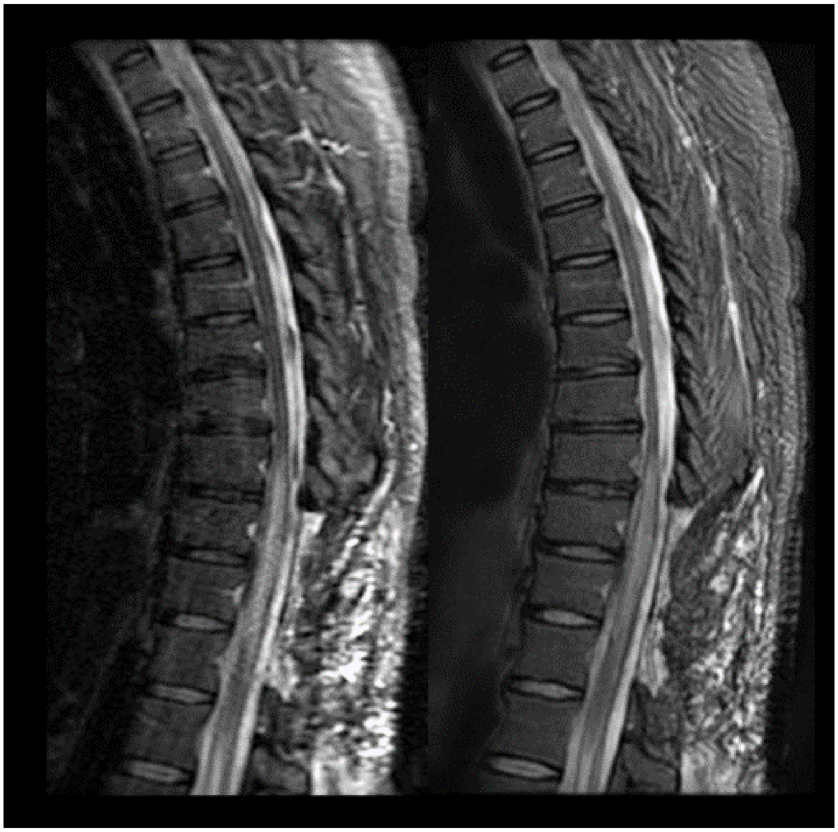

FIG SUPPLEMENTAL: Myelopathy: Conventional acquired STIR (left) and DL synthesized STIR (right). Note the apparent improved SNR and CNR and reduced motion artifact of the synthetic image derived from the high SNR acquired T1w and T2w images.

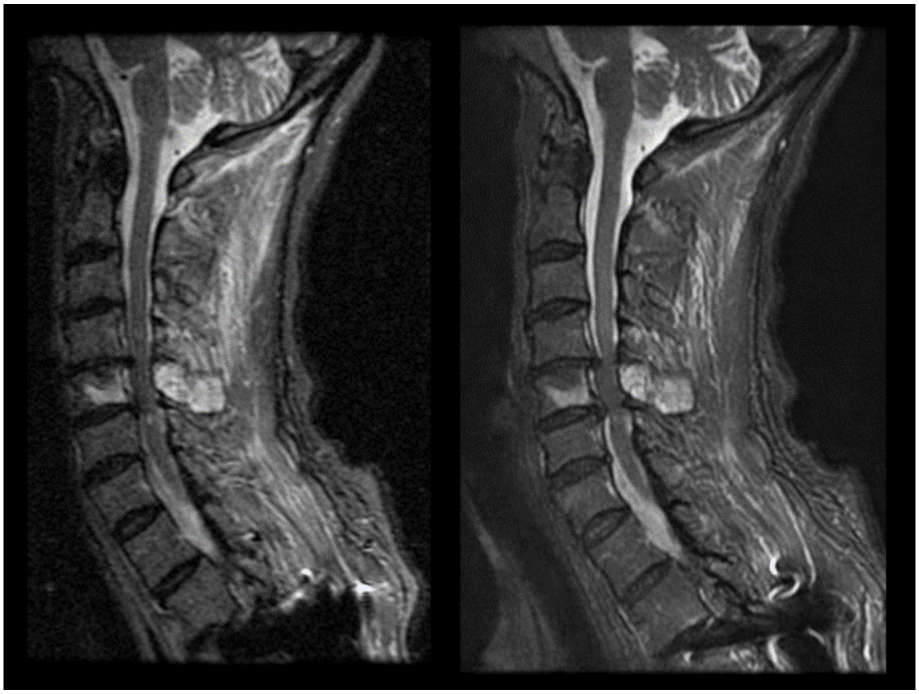

FIG SUPPLEMENTAL: Susceptibility artifact reduction: Acquired STIR (left) and DL synthesized STIR (right). Note the apparent improved SNR,CNR and reduced motion artifact of the synthetic image. Syn-STIR is derived from the high bandwidth T1 and T2 and thus manifests less artifact from hardware than the low bandwidth Acq-STIR.

FIG SUPPLEMENTAL: Osteomyelitis. T1w, T2w and conventional acquired STIR (left) and DL synthesized STIR (far right) images demonstrating excellent concordance between the synthetic and acquired images.

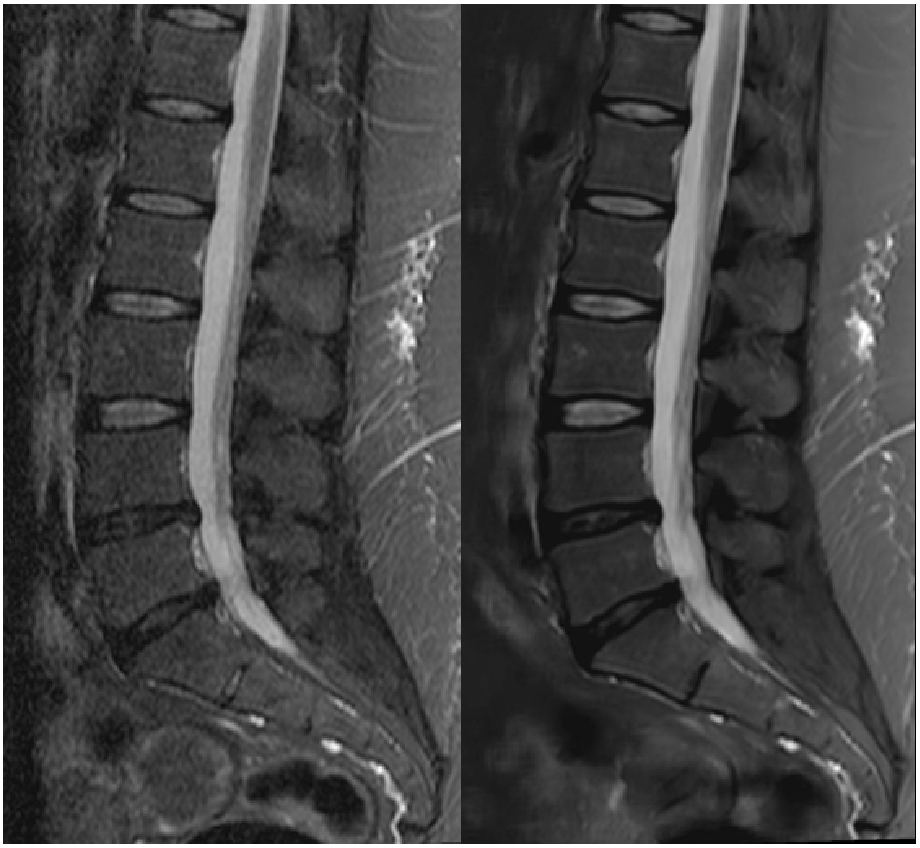

FIG SUPPLEMENTAL: Disc derangements: Conventional acquired STIR (left) and DL synthesized STIR (right). Note the apparent improved SNR and CNR of the synthetic image derived from the high SNR acquired T1w and T2w images

Guidance for readers:

In study 1, you will be asked to classify each case into one of six categories from T1, T2, and STIR images. The six categories are:

1) Degenerative disk disease (DDD)
2) Infection
3) Trauma
4) Cord lesion
5) Non-cord lesion
6) Normal

<u>For multiple categories</u>

- For some cases, they might fall into multiple categories. If so, please feel free to put ALL the types into `comments`, and put one into `diagnosis classification`.

<u>For DDD</u>

- Consider disk bulges, disk dehydration, scoliosis
- If there is DDD with associated degenerative end-plate changes, classify as DDD

<u>For infection</u>

- Standard findings such as bony edema, T2 signal in disk, epidural collections
- Note if there is infection + DDD, classify as infection

<u>For trauma</u>

- Acute/subacute compression fractures
- If Grade 2 or more anterolisthesis with pars defect, classify as trauma
- Note if there is trauma + DDD, classify as trauma

<u>For cord lesion</u>

- If there is cord lesion + DDD, classify as cord lesion
- If there is cord lesion + trauma, classify as cord lesion
- Consider these entities as cord lesions
    - MS/NMOSD/transverse myelitis
    - Intramedullary tumors
    - Chiari malformations
    - Cord infarct
    - Traumatic cord contusion

<u>For non-cord lesions</u>

- If there is non-cord lesion + DDD, classify as non-cord lesion
- If there are bony lesions, even if they extend to the endplates, without significant DDD, classify as non-cord lesion
- If there is DDD with associated degenerative end-plate changes, classify as DDD
- Consider these entities as non-cord lesions:
    - Bone islands
    - Kummel cavities

  - Metastatic lesions
  - Lymphoma involving bone
  - Epidural lipomatosis
  - Arachnoid cyst
  - Schwannoma/neurofibroma
  - Meningioma

For normal

- If none of the diseases applies, please consider it as normal

For image quality, **assess only the STIR image**, using the following scale:

1) Unacceptable
2) Poor
3) Acceptable
4) Good
5) Excellent

In study 2, you will be asked to assess the presence/absence of three entities commonly assessed with STIR for trauma cases. All cases were acquired clinically for an indication of possible traumatic injury. You will only be assessing the STIR images.

The three entities are:

- Pre-vertebral fluid collection (Y/N)
- Bony edema associated with fracture or microtrabecular injury (Y/N)
- Posterior edema/fluid collections associated with soft tissue contusions and posterior ligamentous injury (Y/N)

Cases may have more than one of these entities. Some cases may have none of these entities.

For image quality, **assess only the STIR image**, using the following scale:

1) Unacceptable
2) Poor
3) Acceptable
4) Good
5) Excellent

Stress Testing on Various Noise Levels and Motion Levels

To assess the performance of the model under varying conditions, such as different noise levels or motion levels, a series of simulation tests were conducted on the input data. Specifically, Rician noise [1] with varying levels was added to the T1 and T2 data. Similarly, MR motion[2] with different levels was introduced to the inputs.

Figure S1: Impact of noise levels on T1 and SynthSTIR images. The first row displays input T1 images with varying noise levels (from left to right: reference T1 without added noise, 0.01, 0.02, 0.04, 0.06, and an extreme case of 0.16). The second row shows the corresponding SynthSTIR images, generated by combining the original input and T1 images from the same column.

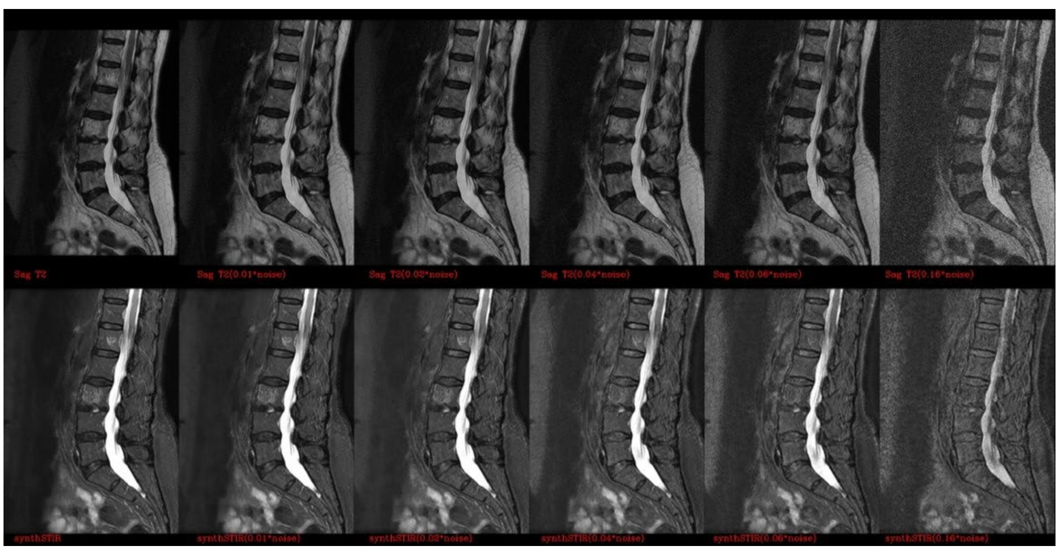

Figure S2: Impact of noise levels on T2 and SynthSTIR images. The first row displays input T2 images with varying noise levels (from left to right: reference T2 without added noise, 0.01, 0.02, 0.04, 0.06, and an extreme case of 0.16). The second row shows the corresponding SynthSTIR images, generated by combining the original input and T2 images from the same column.

According to the observations made in Figure S1 and S2, the image quality of SynthSTIR did not compromise under a reasonable noise level. However, when the noise level for T1 significantly increases, the fat compression performance for the SynthSTIR does not meet expectations. Additionally, when the noise level for T2 is excessively elevated, there is a substantial reduction in the SNR for the SynthSTIR. This means that the signal-to-noise ratio of SynthSTIR images decreases when there is too much noise in the input data.

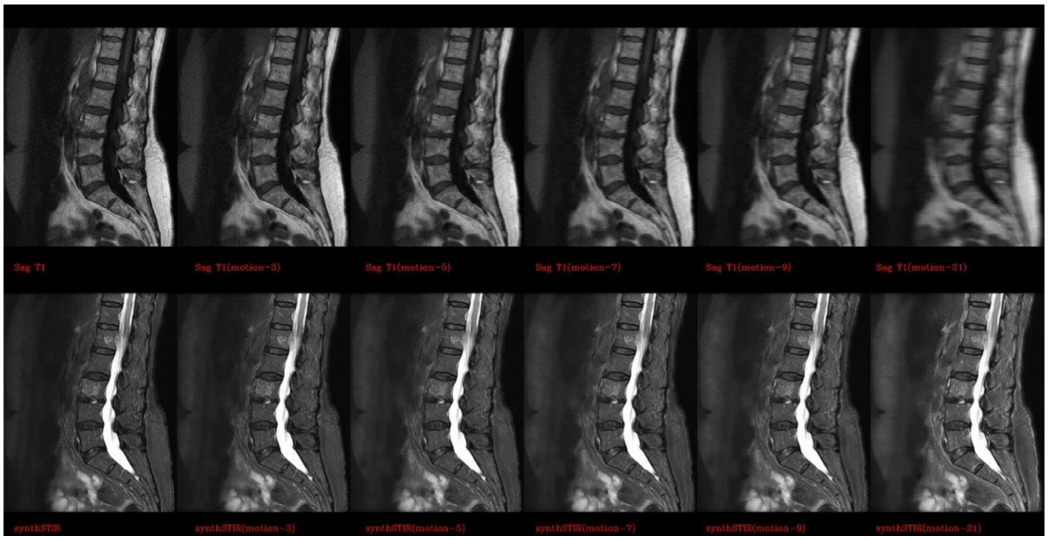

Figure S3: Impact of motion levels on T1 and SynthSTIR images. The first row displays input T1 images with varying motion levels (from left to right: reference T1 without added noise, 3, 5, 7, 9, and an extreme case of 21). The second row shows the corresponding SynthSTIR images, generated by combining the original input and T1 images from the same column.

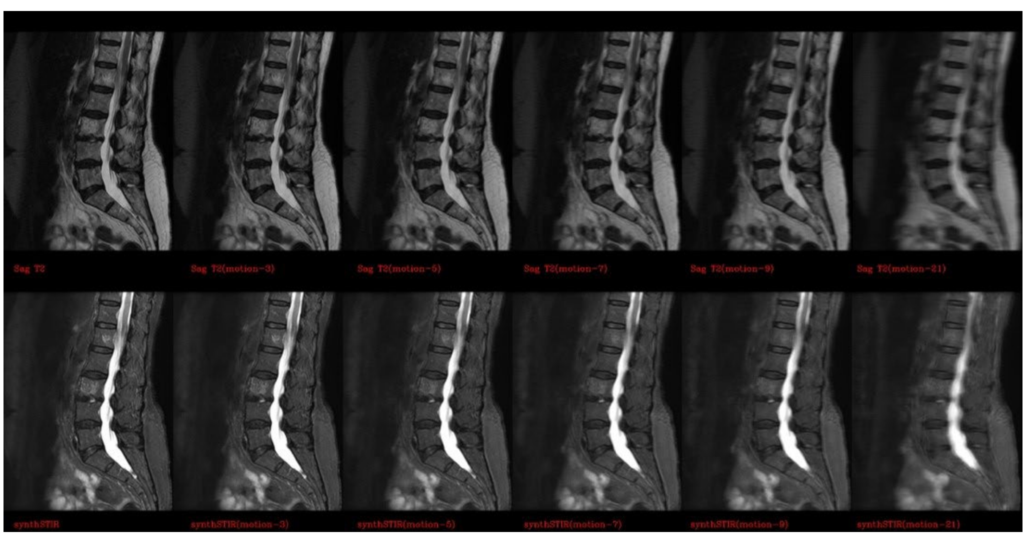

Figure S4: Impact of motion levels on T2 and SynthSTIR images. The first row displays input T2 images with varying motion levels (from left to right: reference T2 without added noise, 3, 5, 7, 9, and an extreme case of 21). The second row shows the corresponding SynthSTIR images, generated by combining the original input and T2 images from the same column.

Based on Figure S3 and S4, minor movements during T1 or T2 imaging do not significantly impact the final synthetic results. However, when the motion is substantial, particularly in T2-weighted images, it can have a noticeable effect on the synthesized images.

Reference:

1. Gudbjartsson, H., & Patz, S. (1995). The Rician distribution of noisy MRI data. American Journal of Neuroradiology, 34(6), 910-914.
2. Perez-Garcia, F., Sparks, R., & Ourselin, S. (2021). TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Computer Methods and Programs in Biomedicine, 106236.

Loss Function Details

When training the neural network to generate synthetic STIR images, the loss function defines the similarity between the acquired STIR images and the synthetic STIR images. For this application we used a weighted combination of multiple loss functions, each designed to teach the network different image properties. These loss functions were L1, SSIM, and Edge. The L1 loss is the absolute difference between the two images on a pixel-by-pixel basis. The SSIM loss is a commonly used image quality metric that assesses the structural similarity between two images [1]. The Edge loss compared the two images after they have been passed through a Sobel filter, which enhances the edges in the image. The Edge loss teaches the network to preserve structural details. Finally, additional L1 losses are calculated using the anatomy and pathology maps to focus the calculation on specific areas of interest within the image.

References

1. A. Horé and D. Ziou, "Image Quality Metrics: PSNR vs. SSIM," 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010, pp. 2366-2369, doi: 10.1109/ICPR.2010.579.