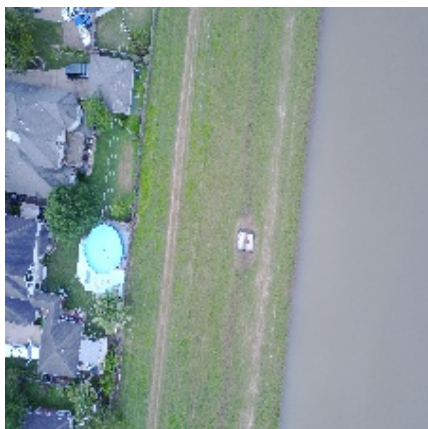| Image-Question pair | Without supervision of visual attention | With supervision of visual attention (Ours) |
| --- | --- | --- |
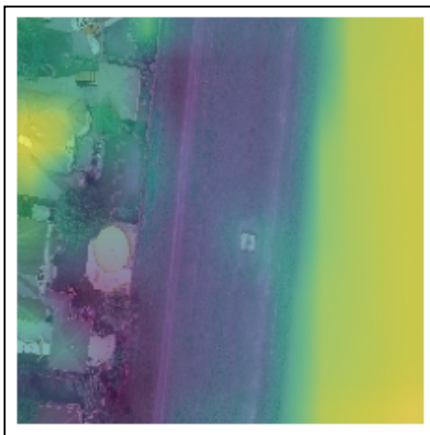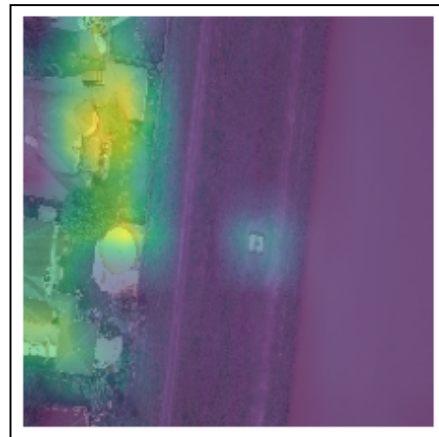


Q: What is the total number of non-flooded buildings?
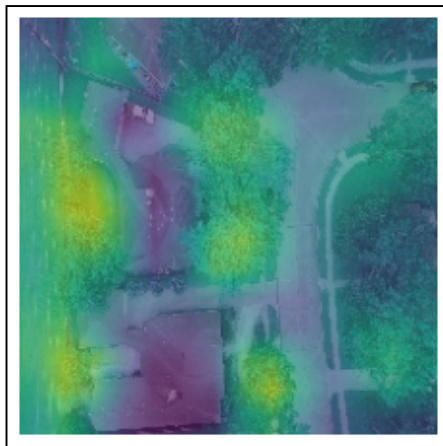
Q: What is the condition of the road?