What Can SPECT Learn from Autoradiography?

Gene Gindi#¶ and Anand Rangarajan§
#Department of Electrical Engineering, SUNY at Stony Brook
§Department of Computer Science, Yale University
¶Department of Radiology, SUNY at Stony Brook

Abstract

For SPECT, where noise and systematic degradations are severe, Bayesian reconstruction appoaches have been advocated for their ability to effectively model the degradations, and to model, through prior density functions, the expected local spatial structure (smoothness) of the class of objects to be reconstructed. These priors are chosen subject to the constraints of mathematical tractability and belief as to the nature of the object. We propose to use autoradiography as a source of "ground truth" functional objects, and show how these may be used as training data to compute a smoothing hyperparameter in a commonly used form of prior in which differences between adjacent pixels are penalized as the sum of the squares of their differences. A discussion of problems in conditioning autoradiographic data for use as ground truth data in SPECT is included, as is a brief desciption of the image formation process in the autoradiography of radiopharmaceuticals. The approach to hyperparameter learning applies to any data, not just autoradiography, deemed representative of the class of objects to be imaged.

I. Autoradiography

Since there is no way to record ground truth functional patterns in humans by which a SPECT imaging scientist may evaluate reconstruction methods, one often resorts to physical or software phantoms. These are typically simple geometrical phantoms designed for the assessment of resolution or contrast performance, or are idealized versions of the expected functional pattern. An opportunity to record ground truth functional patterns, does exist, however, for animal models in the form of autoradiography. Successful reconstruction of ground truth biological data is appealing, but the validity of the extrapolation from animal models, even advanced primates, to humans remains open. Results in our work, however, may be applied to *any* functional phantom, biological or synthetic, that is deemed sufficiently realistic and accurate.

While autoradiography has become a standard research tool in much of biological science, it is less familiar in radiological imaging. Typically, autoradiographs are formed by sacrificing the animal after radiopharmaceutical injection, quick freezing, and sectioning the tissue with a microtome. The thin (20μ) frozen tissue sections are placed in contact with film (see Fig. 1). Common radionuclide SPECT gamma emitters, including Tc-99m, also emit short range β radiation [1] that exposes photographic film. The spatial resolution of the autoradiograph, due primarily to the short range of the β radiation in the tissue and film emulsion [2], [3], is approx 50μ; well below the spatial quantization distance of a film digitizer designed to produce a 256×256 digital autoradiograph at SPECT resolution, as we eventually do. The signal-to-noise ratio (SNR) of the film optical density due to statistical variation in β emission and in film noise effects depends on exposure time, but can be made much lower than the Poisson noise associated with γ-emission. While careful calibration can yield quantitative physiological data from autoradiographs [4], our concern is the accurate mapping of autoradiographic optical density to an accurate value of *relative* radionuclide concentration. To this end, careful calibration to compensate the film transfer character [2] is required, and calibration for the differing β-attenuation properties of grey vs. white vs. CSF brain tissue is required [2] [5]. Finally, adequate compensation for the effects of film digitization [3] must be performed to ensure that the digitized values are reasonably accurate version of relative radionuclide density. To obtain a three dimensional phantom, the thin two-dimensional sections must be aligned and stacked in 3D at an appropriate resolution. Many of the technical problems in this latter stage have, in fact, been addressed by the neuroscience community [6]. (In our work to date, we use 2D autoradiographic data only.) Figure 2 displays a sagittal section of rhesus monkey autoradiograph of the blood flow agent ^{99m}Tc-ECD, and Fig. 3 that of a transverse section of a vervet monkey for Iomazenil (I^{123}), a benzodiazepene neuroreceptor. The detail in these is stunning when viewed from the perspective of noisy and blurry conventional SPECT brain reconstructions.

II. A Bayesian Model

In Bayesian approaches to emission computed tomography, prior information has often taken the form of assumptions on local spatial structure, and many models based on assumptions of simple smoothness or piecewise "flat" (regionally smooth and approximately constant) smoothness have been proposed. Based on our observations of autoradiography, we have proposed a prior model [7] that accounts for piecewise linear regions as well. While simple observation of the autoradiograph may inspire mathematical models that capture certain aspects of smoothness, these models contain free parameters (hyperparameters). Actual performance is dependent on the values of the hyperparameters of these models

The MAP approach in the context of Bayesian framework is to estimate the source field $\mathbf{f}$ by maximizing the posterior probability, given as

$$\Pr(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g}) = \frac{\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})\Pr(\mathbf{F}=\mathbf{f})}{\Pr(\mathbf{G}=\mathbf{g})}, \quad (1)$$

where $\mathbf{f}$ and $\mathbf{g}$ are 2-D vector fields for the source intensities and projection data, respectively, and $\mathbf{F}$ and $\mathbf{G}$ are the associated random fields. Given the posterior distribution in (1), maximizing the posterior distribution is equivalent to minimizing $(-\log)$ of the posterior probability, and the MAP estimation reduces to

$$(\hat{\mathbf{f}}) = \arg\min_{(\mathbf{f})}[-\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) - \log\Pr(\mathbf{F}=\mathbf{f})],$$

where the two terms on the right side are the likelihood and the prior, respectively.

For the likelihood, Poisson statistics are applied in a conventional way:

$$\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) = \prod_{t\theta} \frac{\bar{g}_{t\theta}^{g_{t\theta}}\exp(-\bar{g}_{t\theta})}{g_{t\theta}!} \quad (2)$$

where $\bar{g}_{t\theta} \stackrel{\text{def}}{=} \sum_{ij}\mathcal{H}_{t\theta,ij}f_{i,j}$. In (2), $g_{t\theta}$ is the number of detected counts in the detector bin indexed by t at angle θ, $\bar{g}_{t\theta}$ is the expected number of counts for a particular source $\mathbf{f}$, and $\mathcal{H}_{t\theta,ij}$ is the probability that a photon emitted from source location (i,j) hits detector bin t at angle θ.

We use the familiar Gibbs distribution to model prior information concerning the local smoothness of the image.

$$\Pr(\mathbf{F}=\mathbf{f}) = \frac{1}{Z}\exp(-E_P(\mathbf{f})) \quad (3)$$

where Z is a normalization factor and $E_P(\mathbf{f})$ is the prior energy. We have examined various forms of priors that promote piecewise smoothness [7], but for our present purposes we focus on a commonly used form that seeks to minimize the sum of squared first derivatives:

$$\begin{aligned} E_P(\mathbf{f}) &= \lambda\sum_{ij}\{(f_{ij}-N_{ij})^2 + (f_{ij}-S_{ij})^2 \\ &\quad +(f_{ij}-E_{ij})^2 + (f_{ij}-W_{ij})^2\} \\ &= \lambda\sum_{ij}E(f_{ij}) \end{aligned} \quad (4)$$

where λ is a positive (smoothing) hyperparameter, and the 4 nearest (clique) neighbors ("North, South, East, and West") are $N_{ij} \stackrel{\text{def}}{=} f_{i,j+1}$, $S_{ij} \stackrel{\text{def}}{=} f_{i,j-1}$, $E_{ij} \stackrel{\text{def}}{=} f_{i-1,j}$, and $W_{ij} \stackrel{\text{def}}{=} f_{i+1,j}$. (The term $E(f_{ij})$ is defined in (4) for later convenience.) In (3) the normalizing constant (partition function) Z is

$$Z = \sum_{\{\mathbf{f}\}}\exp[-E_P(\mathbf{f}))] \quad (5)$$

where summation over $\{\mathbf{f}\}$ denotes a sum over all possible vectors $\mathbf{f}$.

The parameter λ controls the degree of smoothness and usually is estimated empirically, yet this process of trial and error precludes the use of MAP reconstruction in a practical setting. We propose that training data, which may be an autoradiograph, may be used to automatically set these types of hyperparameters. We illustrate a closed form solution for the simple quadratic case. By contrast, in conventional approaches e.g. [8], hyperparameters are usually estimated directly from noisy image data.

III. Hyperparameter Estimation

We use a maximum likelihood (ML) approach to estimate the hyperparameter λ. The source $\mathbf{f}$ is now *assumed known*, from autoradiographic or other data. We may then write the ML formulation for estimating the optimal parameter $\hat{\lambda}$:

$$\hat{\lambda} = \arg\min_{\lambda}\{-\log\Pr[\mathbf{F}=\mathbf{f};\lambda]\} \quad (6)$$

where the righthand expression is the -log likelihood. (The negative log converts the maximization problem to a minimization problem.) Note that we are estimating one hyperparameter from a large number (typically 10,000) of data points. Combining (3), (4) and (6) the ML problem becomes

$$\hat{\lambda} = \arg\min\{\lambda\sum_{ij}E(f_{ij}) - \log\sum_{\{\mathbf{f}\}}Z(\lambda)\} \quad (7)$$

The sum involved in computing the partition function is intractable.

To address the problem of the intracablity of Z, we resort to an approximation of the likelihood, the "pseudo-likelihood" [9], given by

$$\begin{aligned} PL(\mathbf{f};\lambda) &= \prod_{ij}\Pr[F_{ij}=f_{ij} \mid \\ &\qquad F_{i,j+1}=N_{ij}, F_{i,j-1}=S_{ij}, \\ &\qquad F_{i+1,j}=E_{ij}, F_{i-1,j}=W_{ij};\lambda] \end{aligned} \quad (8)$$

It turns out [10], that for our form of energy function,

$$\begin{aligned} PL(\mathbf{f};\lambda) &= \exp(-\lambda(E(f_{ij})) \div \\ &\quad \sum_{w}\exp\{-\lambda((w-N_{ij})^2 + (w-S_{ij})^2 \\ &\quad +(w-E_{ij})^2 + (w-W_{ij})^2)\} \end{aligned} \quad (9)$$

where w is a dummy variable that takes on all possible values of a source pixel. We may insert (9) in place of the log likelihood into (6) and get the following maximum pseudolikelihood problem:

$$\begin{aligned}\hat{\lambda} &= \arg\min_{\lambda} \sum_{ij} \Big\{ \lambda E(f_{i,j}) \\ &+ \sum_{ij} \log \Big[\sum_{w} \exp \big\{ -\lambda[(w - N_{ij})^2 \qquad (10) \\ &+ (w - S_{ij})^2 + (w - E_{ij})^2 + (w - W_{ij})^2] \big\} \Big] \Big\}\end{aligned}$$

As shown in the Appendix, we may perform the minimization in (10) above, to get the result:

$$\begin{aligned}\hat{\lambda} &= N/2G \qquad (11) \\ G &= \frac{1}{4} \sum_{ij} [(2f_{ij} - N_{ij} - S_{ij}) + (2f_{ij} - E_{ij} - W_{ij})]^2\end{aligned}$$

and N is the number of nonzero pixels in the image. Interestingly, the combination of derivatives in (11) suggests a continuous version of G:

$$G = (1/4) \int \int (\nabla^2 f)^2 dx dy \qquad (12)$$

Intuitively, one might think that the value of $\hat{\lambda}$ should be dependent upon the amount of noise (total bin counts) in the projection data. Yet the value of $\hat{\lambda}$ was derived from noise*less* training data. The noise level does enter, however, through the scaling of the autoradiograph values $\mathbf{f}$. which are the mean of the Poisson densities for γ emission at each location. With proper scaling of $\mathbf{f}$ to account for the typical number of bin counts, λ depends on noise after all. It is easy to show that $\mathbf{f}$ should be scaled according to

$$f_{ij} = \frac{K}{\sum_{t,\theta} H_{t,\theta,i,j}} f_{ij}^{norm} \qquad (13)$$

where K equals the desired total bin counts and f^{norm} is a version of the autorad normalized so that $\sum_{ij} f_{ij}^{norm} = 1$. In (11), $\hat{\lambda}$ scales with the inverse square of the count level, i.e. more counts means less smoothing as expected. The value of λ also depends on the sampling distance between pixels, so the sampling distance of the training set should match that of the image to be reconstructed. In summary, one first inserts noise level and geometrical information via (13) to obtain a scaled autoradiograph f_{ij} and inserts this into (11) to obtain $\hat{\lambda}$.

We have observed anecdotally that values of $\hat{\lambda}$ calculated as above do indeed yield good reconstructions in the sense that reconstructions obtained using a range of values of λ achieve an approximate minimum in RMSE at $\lambda = \hat{\lambda}$. In these cases, training and reconstruction are accomplished on the same phantom, and values of f are scaled as above.

IV. Appendix

In this appendix, we minimize (10), repeated here for convenience.

$$\begin{aligned}\hat{\lambda} &= \arg\min_{\lambda} \sum_{ij} \Big\{ \lambda E(f_{i,j}) \\ &+ \sum_{ij} \log \Big[\sum_{w} \exp \big\{ -\lambda[(w - N_{ij})^2 \\ &+ (w - S_{ij})^2 + (w - E_{ij})^2 + (w - W_{ij})^2] \big\} \Big] \Big\} (14)\end{aligned}$$

We may complete the square in the sum in [small brackets] in (14) and reexpress it as $\left(\frac{w-\mu}{\sigma}\right)^2 + A$, where $\mu(N,S,E,W) = \frac{N+S+E+W}{4}$, $\sigma = \frac{1}{2}$, and $A(N,S,E,W) \stackrel{\text{def}}{=} N^2 + S^2 + E^2 + W^2 - \frac{(N+S+E+W)^2}{4}$. (In the definitions for A and μ we temporarily drop subscripts i and j on N, S, E, W for convenience.)

Since w may, in principle, take on any real nonnegative value, we replace the sum over w with an integral: $\sum_w \to \int_0^\infty dw$, and the integration over w becomes:

$$\begin{aligned}&\exp\{-\lambda A(N,S,E,W)\} \int_0^\infty \exp -\lambda \left(\frac{w-\mu}{\sigma}\right)^2 dw \\ &= \frac{1}{2}\sqrt{\frac{\pi}{2}} \exp\{-\lambda A(N,S,E,W)\} \qquad (15)\end{aligned}$$

At this point we have evaluated the term in $\Big[$ large brackets $\Big]$ in (14). Application of the $\sum_{ij} \log(\cdot)$ to the right side of (15) results, after dropping terms independent of λ, in the expression

$$-\frac{N}{2} \log \lambda - \lambda \sum_{ij} A(N,S,E,W) \qquad (16)$$

Here, N, the number of pixels in the autoradiograph, does not include zero pixels (i.e. pixels outside the region of support of the image) as these lead to an estimate of λ based partly and incorrectly on the "flat" parts of the image field.

The insertion of (16) and the expression for $E(f_{ij})$ from (4) into (14) results in the following restatement of the minimization problem:

$$\begin{aligned}\hat{\lambda} &= \arg\min_{\lambda} \{ \lambda \sum_{ij} [(f_{ij} - N_{ij})^2 + (f_{ij} - S_{ij})^2 \\ &+ (f_{ij} - E_{ij})^2 + (f_{ij} - W_{ij})^2] \\ &- \frac{N}{2} \log \lambda - \lambda \sum_{ij} A(N_{ij}, S_{ij}, E_{ij}, W_{ij}) \} \\ &= \arg\min_{\lambda} \{ -\frac{N}{2} \log \lambda + \lambda \sum_{ij} G_{ij} \} \qquad (17)\end{aligned}$$

where the term G_{ij} may, after some algebra, be identified as

$$G_{ij} = \frac{1}{4} [(2f_{ij} - N_{ij} - S_{ij}) + (2f_{ij} - E_{ij} - W_{ij})]^2 \qquad (18)$$

Note that in (18), the two terms in (parentheses) may be identified as discrete second derivatives in the horizontal and vertical directions, respectively. We may identify the latter term in (17) as the discrete version of the number:

$$G \equiv \sum_{ij} G_{ij} \rightarrow \frac{1}{4} \iint \left(\nabla^2 f(x,y)\right)^2 dxdy \qquad (19)$$

The minimization problem has been further reduced to

$$\hat{\lambda} = \arg\min_{\lambda}\{-\frac{N}{2}\log\lambda + \lambda G\} = \arg\min_{\lambda}\{PL(\lambda)\} \qquad (20)$$

which may be solved via $\frac{dPL(\lambda)}{d\lambda} = 0$ to get the solution

$$\hat{\lambda} = N/2G \qquad (21)$$

Equation (21) together with the definition in (19) is our solution.

Note that since

$$\frac{d^2PL}{d\lambda^2} = \frac{N}{2\lambda^2} > 0 \qquad (22)$$

our solution for $\hat{\lambda}$ is indeed a minimum, and that since N and G are clearly nonnegative, $\hat{\lambda}$ is also nonnegative as desired.

For a training set that contains much structure i.e. one that is not smooth, then G is large and $\hat{\lambda}$ is correspondingly small, since we don't expect to smooth the data if the ground truth data is itself not smooth. On the other hand, if the training data $\mathbf{f}$ is flat (perfectly smooth), then $\hat{\lambda} = \infty$ and we are entitled to presume a degenerate (perfectly smooth) solution. Curiously, $\hat{\lambda} = \infty$ even if $\mathbf{f}$ is a pure linear ramp, as the pseudolikelihood is unable to distinguish a ramp from a constant. It is, however, sensitive to curvature, i.e. nonzero values of the second derivatives in $\mathbf{f}$ affect $\hat{\lambda}$.

References

[1] L. T. Dillman and F. C. Von der Lage, "Radionuclide Decay Schemes and Nuclear Parameters for Use in Radiation-Dose Estimation (in NM/MIRD Pamphlet No. 10", Technical report, Oak Ridge National Laboratory, Oak Ridge, TN.

[2] J. L. Lear, "Principles of Single and Multiple Radionuclide Autoradiography", In M. E. Phelps, J. C. Mazziotta, and H. R. Schelbert, editors, *Positron Emission Tomography and Autoradiography*, chapter 5, Raven Press, New York, NY, 1986.

[3] D.L. McEachron, editor, *"Functional Mapping in Biology and Medicine: Computer Assisted Autoradiography"*, Experimental Biology and Medicine, Karger, New York, NY, 1986.

[4] J. Lear, R. Ackermann, M. Kameyama, R. Carson, and M. Phelps, "Multiple-radionuclide autoradiography in evaluation of cerebral function", *J. Cereb. Blood Flow Metab.*, 4, pp. 264–269, 1984.

[5] R. C. Walovitch, T. C. Hill, S. T. Garrity, E. H. Cheesman, B. A. Burgess, D. H. O'Leary, A. D. Watson, M. V. Ganey, R. A. Morgan, and S. J. Williams, "Characterization of Technetium-99m-L,L-ECD for Brain Perfusion Imaging, Part 1: Pharmacology of Technetium-99m ECD in Nonhuman Primates", *Journal of Nuclear Medicine*, 30, pp. 1892–1901, 1989.

[6] L. S. Hibbard, J. S. McGlone, D. W. Davis, and R. A. Hawkins, "Three-DImensional Representation and Analysis of Brain Energy Metabolism ", *Science*, 236, pp. 1641–1646, June 1987.

[7] S. J. Lee, A. Rangarajan, and G. Gindi, "Weak Plate Mechanical Models in Bayesian Reconstruction for Emission Tomography", In *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, volume 3, pp. 1533–1537, November 1993.

[8] Z. Zhou, R. M. Leahy, and E. U. Mumcuoglu, "A Comparative Study of the Effects of Using Anatomical Priors in PET Reconstruction", In *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference*, volume 3, pp. 1749–1753, November 1993.

[9] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", *Journal of the Royal Statistical Society, series B*, 36, pp. 192–236, 1974.

[10] S. Geman and C. Graffigne, "Markov random fields image models and their application to computer vision", In A. M. Gleason, editor, *Proc. of the Intl. Congress of Mathematicians 1986*, pp. 1496–1517, American Mathematical Society, Providence, 1987.

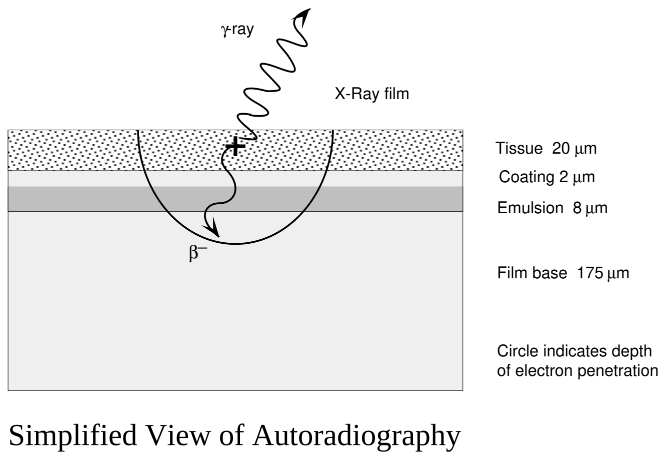

Fig. 1: The emulsion is exposed directly by short range (solid line) β^+ or β^- particles. The γ ray, used in ECT, has little effect on the film exposure due to its long range.

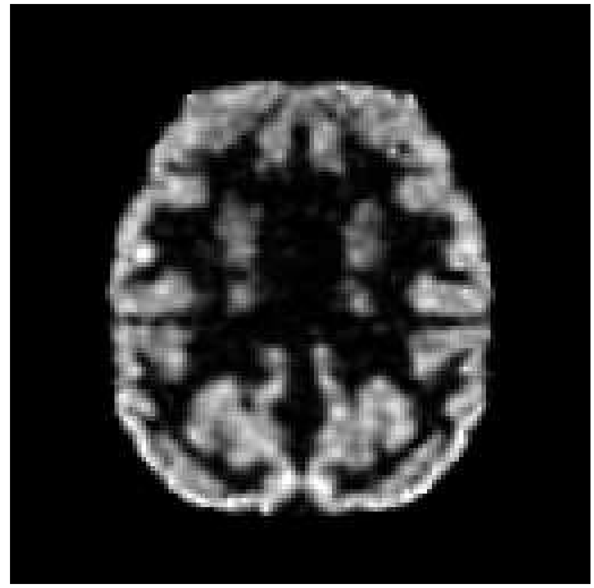

Fig. 4: A reconstruction utilizing the autorad in Fig. 3 as a phantom.

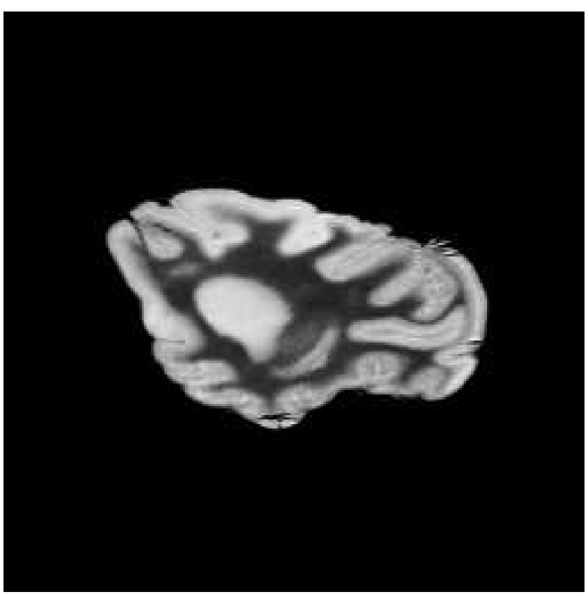

Fig. 2: An example of an autorad obtained from a rhesus. The radiopharmaceutical was ^{99m}Tc-ECD, a blood flow agent. (Courtesy J. Lazewatsky)

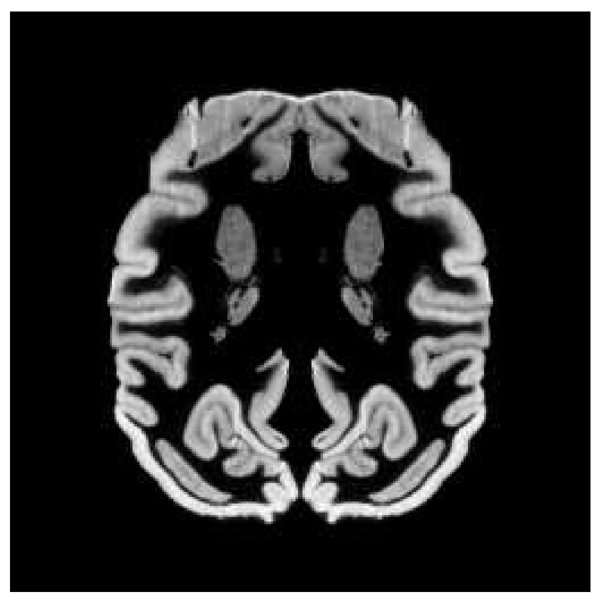

Fig. 3: An example of an autoradiograph obtained with the benzodiazepine neuroreceptor agent Iomazenil (I^{123}). (Courtesy R. Innis)